

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:08:57 ; Search time 1505.66 Seconds  
(without alignments)  
9664.474 Million cell updates/sec  
Title: US-10-083-853b-2\_copy\_1\_500  
Perfect score: 500  
Sequence: 1 gtagatgaagaagcctca.....actaaaggttttcaggct 500

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database 8 GenBank  
1: gb\_batc  
2: gb\_htg:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pl:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sy:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vt:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_mu:  
20: em\_om:  
21: em\_or:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pi:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: em\_vt:  
30: em\_htg\_hum:  
31: em\_htg\_inv:  
32: em\_htg\_other:  
33: em\_htg\_mus:  
34: em\_htg\_pln:  
35: em\_htg\_rod:  
36: em\_htg\_mam:  
37: em\_htg\_vrt:  
38: em\_sy:  
39: em\_htgo\_hum:  
40: em\_htgo\_mus:  
41: em\_htgo\_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	500	100.0	1587	6	AX164746	AX164746 Sequence
C 2	500	100.0	1734	6	AX164761	AX164761 Sequence
C 3	500	100.0	1828	6	AX164744	AX164744 Sequence
C 4	500	100.0	1828	9	AF335584	AF335584 Homo sapi
C 5	500	100.0	2253	9	AF336376	AF336376 Homo sapi
C 6	500	100.0	3718	6	AX365307	AX365307 Sequence
C 7	500	100.0	3736	6	AX365305	AX365305 Sequence
C 8	500	100.0	4070	9	BC030645	BC030645 Homo sapi
C 9	500	100.0	160345	2	AC067870	AC067870 Homo sapi
C 10	500	100.0	180648	9	AP003043	AP003043 Homo sapi
C 11	500	100.0	182403	2	AC024052	AC024052 Homo sapi
C 12	496.8	99.4	3710	9	AY027518	AY027518 Homo sapi
C 13	496.8	99.4	3729	9	AY027517	AY027517 Homo sapi
C 14	482.8	96.6	1882	6	AX044487	AX044487 Sequence
C 15	482.8	96.6	1882	6	AX207483	AX207483 Sequence
C 16	460	92.0	485	6	AX331460	AX331460 Sequence
C 17	133.4	26.7	1428	9	AB033832	AB033832 Homo sapi
C 18	108.6	21.7	2870	10	BC030896	BC030896 Mus muscu
C 19	74.2	14.8	1472	6	AX044538	AX044538 Sequence
C 20	74.2	14.8	1472	6	AX207485	AX207485 Sequence
C 21	71	14.2	161951	2	CNS071PU	AL713931 Oryza sat
C 22	64.8	13.0	253305	3	PFMAL3P7	AL034559 Plasmodiu
C 23	64.2	12.8	97643	9	AL157783	AL157783 Human DNA
C 24	64.2	12.8	171606	2	AL591118	AL591118 Homo sapi
C 25	62.8	12.6	176639	9	AC016987	AC016987 Homo sapi
C 26	62.8	12.6	182430	2	AC104589	AC104589 Homo sapi
C 27	62.8	12.6	189791	2	AC011829	AC011829 Homo sapi
C 28	62.6	12.5	40752	9	AC087065	AC087065 Homo sapi
C 29	62.2	12.4	62707	2	AC091197	AC091197 Homo sapi
C 30	62.2	12.4	186836	9	AC079340	AC079340 Homo sapi
C 31	61.6	12.3	191119	9	AC009957	AC009957 Homo sapi
C 32	61.4	12.3	256	6	AX302693	AX302693 Sequence
C 33	60.6	12.1	148202	2	AL772396	AL772396 danio rer
C 34	60.2	12.0	253305	3	PFMAL3P7	AL034559 Plasmodiu
C 35	59.4	11.9	128769	2	AL139261	AL139261 Homo sapi
C 36	59.2	11.8	171317	9	AC020941	AC020941 Homo sapi
C 37	58.8	11.8	175928	2	AC129750	AC129750 Rattus no
C 38	58.8	11.8	192177	2	AC109118	AC109118 Rattus no
C 39	58.2	11.6	128234	2	AC090192	AC090192 Homo sapi
C 40	58.2	11.6	188971	9	AC007707	AC007707 Homo sapi
C 41	58	11.6	151498	9	AC099331	AC099331 Homo sapi
C 42	57.4	11.5	162075	9	HS127D3	AL021026 Homo sapi
C 43	56.8	11.4	175103	9	AC109135	AC109135 Homo sapi
C 44	56.6	11.3	61306	9	AC007423	AC007423 Homo sapi
C 45	56.6	11.3	162401	2	AC024024	AC024024 Homo sapi

ALIGNMENTS

RESULT 1  
AX164746/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AX164746  
Sequence 3 from Patent WO0125437.  
AX164746  
AX164746.1 GI:14545599  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1587)  
Shimkets,R.A., Lichenstein,H., Herrmann,J.L., Boldog,F.L.,  
Minskoff,S. and Jeffers,M.  
Growth factor polypeptides and nucleic acids encoding same

AX164746 1587 bp DNA linear PAT 22-JUN-2001







```

RESULT 6
AX365307/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

```

BASE COUNT	sig_peptide	1157 a	730 c	715 g	1116 t	
ORIGIN						
	Query Match	100.0%;	Score 500;	DB 6;	Length 3718;	
	Best Local Similarity	100.0%;	Pred. No. 5.1e-83;			
	Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	GTATATGTAAGAAGCCTCATCTTTTGATTTTTAAATATACAAGATGCTTTCTTTAAAGAGA	60			
Db	1732	GTATATGTAAGAAGCCTCATCTTTTGATTTTTAAATATACAAGATGCTTTCTTTAAAGAGA	1673			
QY	61	GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAAAATAAATTTATCTCCTAAATTTT	120			
Db	1672	GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAAAATAAATTTATCTCCTAAATTTT	1613			
QY	121	CTAAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT	180			
Db	1612	CTAAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT	1553			
QY	181	AACTCAAACATATGTAAGCTCTGGTCTACTGCTGTTATATATACCAAAAAAACAATTTGAT	240			
Db	1552	AACTCAAACATATGTAAGCTCTGGTCTACTGCTGTTATATATACCAAAAAAACAATTTGAT	1493			
QY	241	CTATATACACATAGACATGAATATATTCTGTGTGTGTTTGTGCATATATAACCTCAAAC	300			
Db	1492	CTATATACACATAGACATGAATATATTCTGTGTGTGTTTGTGCATATATAACCTCAAAC	1433			
QY	301	ACTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATACCTTTCTAC	360			
Db	1432	ACTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATACCTTTCTAC	1373			
QY	361	TTGCCATGCGATTAAACAAAGACGCTGAGCTCAGCAACCACTTGTGTTTCATTGCATTG	420			
Db	1372	TTGCCATGCGATTAAACAAAGACGCTGAGCTCAGCAACCACTTGTGTTTCATTGCATTG	1313			
QY	421	CAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAAAGGGTCTCTTATCTCACCTCCTTAA	480			

Db 1312 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAAGGGTCTCTATCTCACCCCTCCTAA 1253

```
Qy 481 ACTAAGGTTCTTTCAGGCT 500
Db 1252 ACTAAGGTTCTTTCAGGCT 1233

RESULT 7
AX365305/c 3736 bp DNA linear PAT 15-FEB-2002
LOCUS
DEFINITION Sequence 1 from Patent W00189450.
ACCESSION AX365305
VERSION AX365305.1 GI:18697036
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Beals, J.M., Gonzalez-Dewhitt, P.A., Hammond, L.J., Lu, J., Na, S.,
Su, E.W., Wicher, D.R. and Wroblewski, V.J.
"Breathing musculoskeletal disorders using lp85 and analogs thereof
(Patent No. 01/09450-A) 1-29-NOV-2001."
JOURNAL OF PATENT AND COMPANY (US)
FEATURES
Location/Qualifiers
source 1..3736
/organism="Homo sapiens"
/db_xref="taxon:9606"
114..1226
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD23805.1"
/db_xref="GI:18697037"
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRREDSE
HLRDLYRDETIQKNGYVQSPFPNSYPRNLLTWRLHSEQNTRIQLVFDNFGLE
EANDICRDEVEDISESTIIRGWGCHKVPPRIKSRNQIKITPKSDDYFVAK
PGKIVYSLLEDQPAASSETNWSVTSSISGYSNPSVDTPTLADALDKKIAEFD
TVEDLKYPNWSQEDLENMIDTPRGRSHDRKSKVDLDRLNDKARKYRSCYPRN
YSNIREELKLANVFFPCLLVQRCGNGCGTWNWRSCTCNSGKTVKKYHEVLQFE
PGHKRGRKTMALVDIQLDHERDCICSSRPPR"
sig_peptide 114..149
BASE COUNT 1164 a 736 c 718 g 1118 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 3736;
Best Local Similarity 100.0%; Pred. No. 5,1e-83;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTATATGAGAAGCCTCATCTTTGATTTTAAATATACAAAGATGCTTCTTAAAGAGA 60
Db 1750 GTATATGAGAAGCCTCATCTTTGATTTTAAATATACAAAGATGCTTCTTAAAGAGA 1691

Qy 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 120
Db 1690 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 1631

Qy 121 CTAAGACATGTTTATATATTTGACATCCCTTATTTTGGCAAGGATTTTAAAGATCT 180
Db 1630 CTAAGACATGTTTATATATTTGACATCCCTTATTTTGGCAAGGATTTTAAAGATCT 1571

Qy 181 AACTCAACATATATGATGCTCTGGTGTACCTGGTTATATATACAAAATAATTTGAT 240
Db 1570 AACTCAACATATATGATGCTCTGGTGTACCTGGTTATATATACAAAATAATTTGAT 1511

Qy 241 CTATATACATACATGATGAATATATTTCTGTGTGTTTGTGCATATATAACCTCAAC 300
Db 1510 CTATATACATACATGATGAATATATTTCTGTGTGTTTGTGCATATATAACCTCAAC 1451

Qy 301 ACTATTATTAAATGCAATCCTATATTTCTTAAAGTATAGAAGTTGATGATATACCTTTCTAC 360
Db 1450 ACTATTATTAAATGCAATCCTATATTTCTTAAAGTATAGAAGTTGATGATATACCTTTCTAC 1391

Qy 361 TTGCCATGGCATTAACAAAGCAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATTG 420
Db 1390 TTGCCATGGCATTAACAAAGCAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATTG 1331
```

```
Qy 421 CAGGCTAGTAGTAAGTTTGTCTGTAGAAAAGGCTCTTATCTACCCCTCTAA 480
Db 1330 CAGGCTAGTAGTAAGTTTGTCTGTAGAAAAGGCTCTTATCTACCCCTCTAA 1271

RESULT 8
BC030645/c 4070 bp mRNA linear PRI 21-MAY-2002
LOCUS
DEFINITION Homo sapiens, spinal cord-derived growth factor-B, clone MGC:36867
IMAGE:4824526, mRNA, complete cds.
ACCESSION BC030645
VERSION BC030645.1 GI:21040416
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 4070)
Strausberg, R.
Direct Submission
Submitted (20-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-help@nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 33 Row: 1 Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 15451919.
FEATURES
Location/Qualifiers
source 1..4070
/organism="Homo sapiens"
/db_xref="LocusID:80310"
/db_xref="taxon:9606"
/clone="MGC:26867 IMAGE:4824526"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_post="DH10B"
/notes="Vector: pBluescript"
458..1552
/codon_start=1
/product="spinal cord-derived growth factor-B"
/protein_id="AAH30645.1"
/db_xref="GI:21040417"
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRDDLY
RDDETQKNGYVQSPFPNSYPRNLLTWRLHSEQNTRIQLVFDNFGLEAENDI
CRYDFVEDISESTIIRGWGCHKVPPRIKSRNQIKITPKSDDYFVAKPGFKIY
YSLLEDQPAASSETNWSVTSSISGYSNPSVDTPTLADALDKKIAEFDVLDL
KYPNWSQEDLENMIDTPRGRSHDRKSKVDLDRLNDKARKYRSCYPRNYSNIR
EELKLANVFFPCLLVQRCGNGCGTWNWRSCTCNSGKTVKKYHEVLQFEFGHKR
RGRKTMALVDIQLDHERDCICSSRPPR"
BASE COUNT 1227 a 843 c 831 g 1169 t
ORIGIN

Query Match 100.0%; Score 500; DB 9; Length 4070;
```

\* 110141 130770: contig of 20630 bp in length

★ 91903 110040: contig of 18138 bp in length  
★ 110041 110140: gap of 100 bp  
★ 110141 130770: contig of 20630 bp in length

\* 130771 130870: gap of 100 bp  
 \* 130871 160345: Contig of 29475 bp in length.  
 FEATURES  
 source

Location/Qualifiers  
 1. .160345  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11"

/clone="RP11-475J2"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature  
 1. .946  
 /note="assembly\_fragment"

clone\_end:r7  
 vector\_side:right

misc\_feature  
 1047. .2803  
 /note="assembly\_fragment"

misc\_feature  
 2904. .4474  
 /note="assembly\_fragment"

misc\_feature  
 4575. .8112  
 /note="assembly\_fragment"

misc\_feature  
 8213. .11442  
 /note="assembly\_fragment"

misc\_feature  
 11543. .14027  
 /note="assembly\_fragment"

misc\_feature  
 14128. .17482  
 /note="assembly\_fragment"

misc\_feature  
 17583. .20007  
 /note="assembly\_fragment"

clone\_end:SP6  
 vector\_side:right

misc\_feature  
 20108. .26548  
 /note="assembly\_fragment"

misc\_feature  
 26649. .34423  
 /note="assembly\_fragment"

misc\_feature  
 34524. .41559  
 /note="assembly\_fragment"

misc\_feature  
 41660. .50886  
 /note="assembly\_fragment"

misc\_feature  
 50987. .64203  
 /note="assembly\_fragment"

misc\_feature  
 64304. .75627  
 /note="assembly\_fragment"

misc\_feature  
 75728. .91802  
 /note="assembly\_fragment"

misc\_feature  
 91903. .110040  
 /note="assembly\_fragment"

misc\_feature  
 110141. .130770  
 /note="assembly\_fragment"

misc\_feature  
 130871. .160345  
 /note="assembly\_fragment"

BASE COUNT 48193 a .30956 c 30439 g 49055 t 1702 others

ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 160345;

Best Local Similarity 100.0%; Pred. No. 2e-83;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAGAAGCCTCATCTTTGATTTTTAAATATACAAAGTCTTTCTTTAAAGAGA 60

DB 82195 GTATATGTAGAAGCCTCATCTTTGATTTTTAAATATACAAAGTCTTTCTTTAAAGAGA 82136

QY 61 GCAAGATTCAAAATCTTTTGTCTTCAAAATTTAAATTTATCTCTCTAAATTTT 120

DB 82135 GCAAGATTCAAAATCTTTTGTCTTCAAAATTTAAATTTATCTCTCTAAATTTT 82076

QY 121 CTAAGACATGTTTCATATATTGACACCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 180

DB 82075 CTAAGACATGTTTCATATATTGACACCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 82016

QY 181 AACTCAACATATGTAGCTCTGGTCTACCTGTTATATATACCAAAAAAAGATTGTAT 240

DB 82015 AACTCAACATATGTAGCTCTGGTCTACCTGTTATATATACCAAAAAAAGATTGTAT 81956

QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAAC 300

DB 81955 CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAAAC 81896

QY 301 ACTATTATTAAATGCAATCTTATATTTCTTAGGTATAGAAGTTCATATACCTTTCTAC 360

DB 81895 ACTATTATTAAATGCAATCTTATATTTCTTAGGTATAGAAGTTCATATACCTTTCTAC 81836

QY 361 TTGCCATGCGATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGCAATG 420

DB 81835 TTGCCATGCGATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGCAATG 81776

QY 421 CAGGCTAGTAGTAAGTTTGTCTGGTAGGAAAGGGTCTCTTATCTCACCCCTCTTAA 480

DB 81775 CAGGCTAGTAGTAAGTTTGTCTGGTAGGAAAGGGTCTCTTATCTCACCCCTCTTAA 81716

QY 481 ACTAAAGGTTCTTTTCAGGCT 500

DB 81715 ACTAAAGGTTCTTTTCAGGCT 81696

RESULT 10

AP003043

LOCUS

DEFINITION

AP003043

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .180648

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="11g"

/clone="RP11-617B3"

BASE COUNT 56234 a 35375 c 33753 g 55286 t

ORIGIN

Query Match 100.0%; Score 500; DB 9; Length 180648;

Best Local Similarity 100.0%; Pred. No. 1.9e-83;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAGAAGCCTCATCTTTGATTTTTAAATATACAAAGTCTTTCTTTAAAGAGA 60

DB 39863 GTATATGTAGAAGCCTCATCTTTGATTTTTAAATATACAAAGTCTTTCTTTAAAGAGA 39922

QY 61 GCAAGATTCAAAATCTTTTGTCTTCAAAATTTAAATTTATCTCTCTAAATTTT 120

DB 39923 GCAAGATTCAAAATCTTTTGTCTTCAAAATTTAAATTTATCTCTCTAAATTTT 39982

QY 121 CTAAGACATGTTTCATATATTGACACCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 180

DB 39983 CTAAGACATGTTTCATATATTGACACCCCTTTATTTGGCAAGAGATTTTAAGAGTCT 40042

QY 181 AACTCAACATATGTAAGCTCTGGTACTGCTTATATATACCAAAAAACATTTGAT 240  
 |||||  
 Db 40043 AACTCAACATATGTAAGCTCTGGTACTGCTTATATATACCAAAAAACATTTGAT 40102  
 |||||  
 QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 300  
 |||||  
 Db 40103 CTATATACATAGACATGAATATATTTCTGTGTGTTGTCATATATACCTCAAC 40162  
 |||||  
 QY 301 ACTATTATTAATGAATCCTATATCTTTAGCTATAGAAAGTTGATGATATACCTTTCTAC 360  
 |||||  
 Db 40163 ACTATTATTAATGAATCCTATATCTTTAGCTATAGAAAGTTGATGATATACCTTTCTAC 40222  
 |||||  
 QY 361 TTGCCATGCATTAACAAGCAAGGCTGAGACTAGCAACACCTTTGTTCATTGCATTG 420  
 |||||  
 Db 40223 TTGCCATGCATTAACAAGCAAGGCTGAGACTAGCAACACCTTTGTTCATTGCATTG 40282  
 |||||  
 QY 421 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGGTCTTATATCACCCTCTTAA 480  
 |||||  
 Db 40283 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGGTCTTATATCACCCTCTTAA 40342  
 |||||  
 QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
 |||||  
 Db 40343 ACTAAAGGTTCTTTTCAGGCT 40362  
 |||||

RESULT 11  
 AC024052  
 LOCUS  
 DEFINITION Homo sapiens chromosome 11 clone RP11-617B3, WORKING DRAFT  
 SEQUENCE, 7 unordered pieces.  
 AC024052  
 AC024052.3 GI:9838295  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens.  
 SOURCE Homo sapiens.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 182403)  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 AUTHORS Waterston,R.H.  
 REFERENCE 2 (bases 1 to 182403)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
 COMMENT On Aug 17, 2000 this sequence version replaced gi:7109658.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H.NH0617803  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Chemistry: Dye-terminator Big Dye; 0%  
 Chemistry: Dye-terminator Big Dye; 0%  
 Chemistry: Dye-terminator Big Dye; 0%  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 178574 bases at least Q40  
 Consensus quality: 179480 bases at least Q30  
 Consensus quality: 179940 bases at least Q20  
 Insert size: 198000; agarose-fp  
 Insert size: 183143; sum-of-ctg  
 Quality coverage: 5.39 in Q20 bases; agarose-fp  
 Quality coverage: 5.87 in Q20 bases; sum-of-ctg

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1666: contig of 1666 bp in length  
 \* 1667: gap of unknown length  
 \* 1767: contig of 7367 bp in length  
 \* 9133: gap of unknown length  
 \* 9234: contig of 11430 bp in length  
 \* 20663: gap of unknown length  
 \* 20664: contig of 15164 bp in length  
 \* 20764: gap of unknown length  
 \* 35927: contig of 25583 bp in length  
 \* 35928: gap of unknown length  
 \* 36028: contig of 34457 bp in length  
 \* 61611: gap of unknown length  
 \* 61710: contig of 34457 bp in length  
 \* 61711: gap of unknown length  
 \* 96168: contig of 86136 bp in length.  
 \* 96268: contig of 86136 bp in length.

## FEATURES

source  
 1..182403

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"

/clone="RP11-617B3"  
 1..1666

misc\_feature

/note="assembly\_name:Contig2"  
 1767..9133

misc\_feature

/note="assembly\_name:Contig3"  
 clone\_end:SP6  
 vector\_side:left

misc\_feature

9234..20663  
 /note="assembly\_name:Contig4"

misc\_feature

20764..35927  
 /note="assembly\_name:Contig5"

misc\_feature

36028..61610  
 /note="assembly\_name:Contig6"

misc\_feature

61711..96167  
 /note="assembly\_name:Contig7"

misc\_feature

96268..182403  
 /note="assembly\_name:Contig8"  
 clone\_end:T7  
 vector\_side:left

BASE COUNT 55923 a 35190 c 34519 g 56169 t 602 others  
 ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 182403;

Best Local Similarity 100.0%; Pred. No. 1.9e-83;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGGATTTTAAATATACAGATGCTTTCTTTAAGAGA 60

Db 136115 GTATATGTAAGAAAGCCCTCATCTTTTGGATTTTAAATATACAGATGCTTTCTTTAAGAGA 136174

QY 61 GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 120

Db 136175 GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 136234

QY 121 CTAAGACATGTTTCATATATTTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180

Db 136235 CTAAGACATGTTTCATATATTTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 136294

QY 181 AACTCAACATATGTAAGCTCTGGTGTACCTGGTGTATATATACCAAAAAACATTTGAT 240

Db 136295 AACTCAACATATGTAAGCTCTGGTGTACCTGGTGTATATATACCAAAAAACATTTGAT 136354

QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATACCTCAAC 300

Db 136355 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATACCTCAAC 136414

QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTGATGATATACCTTTCTAC 360

Db 136415 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTGATGATATACCTTTCTAC 136474



QY	361	TTGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	420	QY	61	GCAAGATTCAAAATTTGTTTGTTCAAAAATTTAAAAATAAATTTATCTCCTAAATTTT	120
Db	136475	TTGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	136534	Db	1651	GCAAGATTCAAAATTTGTTTGTTCAAAAATTTAAAAATAAATTTATCTCCTAAATTTT	1592
QY	421	CAGGCTAGTAGTAAGTGGTGGTGGTAGGAAAAGGCTCTTATCTACCTCCTTAA	480	QY	121	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAGAGTCT	180
Db	136535	CAGGCTAGTAGTAAGTGGTGGTGGTAGGAAAAGGCTCTTATCTACCTCCTTAA	136594	Db	1591	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAGAGTCT	1532
QY	481	ACTAAAGTTCCTTCAGGCT 500		QY	181	AACCAACATATGTAAGCTCTGGTCTACCTGGTATATATACCAAAAAACATTTGAT	240
Db	136595	ACTAAAGTTCCTTCAGGCT 136614		Db	1531	AACCAACATATGTAAGCTCTGGTCTACCTGGTATATATACCAAAAAACATTTGAT	1472
RESULT 12				QY	241	CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAC	300
LOCUS	AY027518/c		3710 bp mRNA linear PRI 17-JUL-2002	Db	1471	CTATATACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAAC	1412
DEFINITION	Homo sapiens iris-expressed growth factor short form (IEGF) mRNA, complete cds, alternatively spliced.			QY	301	ACTATATTAATGAATCCCTATATTTCTAGGTATAGAAAGTGTGATATACCTTCTAC	360
ACCESSION	AY027518			Db	1411	ACTATATTAATGAATCCCTATATTTCTAGGTATAGAAAGTGTGATATACCTTCTAC	1352
VERSION	AY027518.1	GI:13432062		QY	361	TTGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	420
KEYWORDS	human.			Db	1351	TTGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCATTCGATG	1292
SOURCE	human.			QY	421	CAGGCTAGTAGTAAGTGGTGGTGGTAGGAAAAGGCTCTTATCTCCTCCTTAA	480
ORGANISM	Homo sapiens			Db	1291	CAGGCTAGTAGTAAGTGGTGGTGGTAGGAAAAGGCTCTTATCTCCTCCTTAA	1232
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			QY	481	ACTAAAGTTCCTTCAGGCT 500	
AUTHORS	Wistow,G., Berstein,S.L., Ray,S., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K.			Db	1231	ACTAAAGTTCCTTCAGGCT 1212	
TITLE	Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium			RESULT 13			
JOURNAL	Mol. Vis. 8, 185-195 (2002)			AY027517/c			
MEDLINE	22103462			LOCUS	AY027517	3729 bp mRNA linear PRI 17-JUL-2002	
PUBMED	12107412			DEFINITION	Homo sapiens iris-expressed growth factor long form (IEGF) mRNA, complete cds, alternatively spliced.		
REFERENCE	2 (bases 1 to 3710)			ACCESSION	AY027517		
AUTHORS	Wistow,G.			VERSION	AY027517.1	GI:13432060	
TITLE	Direct Submission			KEYWORDS	human.		
JOURNAL	Submitted (13-FEB-2001) MSF, NEI, 6/331, NIH, Bethesda, MD 20892-2740, USA			SOURCE	Homo sapiens		
FEATURES	Location/Qualifiers			ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
BASE COUNT	1169 a 724 c 703 g 1114 t			REFERENCE	1 (bases 1 to 3729)		
ORIGIN				AUTHORS	Wistow,G., Berstein,S.L., Ray,S., Wyatt,M.K., Behal,A., Touchman,J.W., Bouffard,G., Smith,D., and Peterson,K.		
Query Match	99.4%; Score 496.8; DB 9; Length 3710;			TITLE	Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium		
Best Local Similarity	99.6%; Pred. No. 2e-82;			JOURNAL	Mol. Vis. 8, 185-195 (2002)		
Matches 498; Conservative	0; Mismatches 2; Indels 0; Gaps 0;			MEDLINE	22103462		
QY	1	GATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA	60	PUBMED	12107412		
Db	1711	GATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA	1652	REFERENCE	2 (bases 1 to 3729)		
				AUTHORS	Wistow,G.		
				TITLE	Direct Submission		
				JOURNAL	Submitted (13-FEB-2001) MSF, NEI, 6/331, NIH, Bethesda, MD 20892-2740, USA		
				FEATURES	Location/Qualifiers		
				BASE COUNT	1169 a 724 c 703 g 1114 t		
				ORIGIN			
				Query Match	99.4%; Score 496.8; DB 9; Length 3710;		
				Best Local Similarity	99.6%; Pred. No. 2e-82;		
				Matches 498; Conservative	0; Mismatches 2; Indels 0; Gaps 0;		
				QY	1	GATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA	60
				Db	1711	GATATGTAAGAAAGCCCTCATCTTTTGTATTTTATATACAGATGCTTCTTTAAGAGA	1652



/protein\_id="AAK20081.1"  
/db\_xref="GI:13432061"  
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRDEN  
HLTDLVRRDTIQVKGNGYQSPFNPSYPRNLLTWRHLSQENTRIQLVFDNOFGLE  
EAENDICRYDEVEVEDISESTIIRGRCGKHKEVPPRIKSRNQIKITFKSDDYFVAK  
PGKIYVSLLEDFQPAASSETNWSVTSSISGVSYNSPVDPTLIADALDKKIAEFD  
TVEDLLKYPNPSQEDLENNYLDTPRYGRSYHDKRSKVDLDRLNDADKRYSCPTPN  
YSNIREELKLANVFFPRCLLVQRCGNGCGCTVNNRSCNCSGKTVKRYHEVLOFE  
PGHKRRGRAKTALVDIQLDHHERCDCICSSRPR"  
BASE COUNT 1177 a 730 c 706 g 1116 t  
ORIGIN

Query Match 99.4%; Score 496.8; DB 9; Length 3729;  
Best Local Similarity 99.6%; Pred. No. 2e-82;  
Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
Db 1729 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1670  
Qy 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAAATTTAAAAATTTAAATTTATCTCCTAAATTTT 120  
Db 1669 GCAAGATTCAAAATTTGTTTGTGTTTCAAAAATTTAAAAATTTAAATTTATCTCCTAAATTTT 1610  
Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGTTTAAAGAGTCT 180  
Db 1609 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGTTTAAAGAGTCT 1550  
Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTGTATATATACCAAAAACATTTGAT 240  
Db 1549 AACTCAACATATGTAAGCTCTGCTGACCTGGTGTATATATACCAAAAACATTTGAT 1490  
Qy 241 CTATATACATAGACATGAATATATTTCTGTGTCTTTGTGTCATATATAACCTCAAAC 300  
Db 1489 CTATATACATAGACATGAATATATTTCTGTGTCTTTGTGTCATATATAACCTCAAAC 1430  
Qy 301 ACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTTCTAC 360  
Db 1429 ACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTTCTAC 1370  
Qy 361 TTGCATGGCATTAACAAGCAAGCTGAGACTCAGCAACCACTTGTTGTCATTGCATTG 420  
Db 1369 TTGCATGGCTTTAACAAGCAAGCTGAGACTCAGCAACCACTTGTTGTCATTGCATTG 1310  
Qy 421 CAGCTAGTAGTAGTTGGTTGCTGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA 480  
Db 1309 CAGCTAGTAGTAGTTGGTTGCTGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA 1250  
Qy 481 ACTAAGGTTCTTTCAGGCT 500  
Db 1249 ACTAAGGTTCTTTCAGGCT 1230

RESULT 14  
AX044487/c  
LOCUS AX044487 1882 bp DNA linear PAT 24-NOV-2000  
DEFINITION Sequence 1 from Patent WO0066736.  
ACCESSION AX044487  
VERSION AX044487.1 GI:11343345  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Gilbert,T., Hart,C.E., Sheppard,P.O. and Gilbertson,D.G.  
TITLE Growth factor homolog zvegfr4  
JOURNAL Patent: WO 0066736-A 1 09-NOV-2000;  
ZymoGenetics, Inc. (US)  
FEATURES  
source  
1. .1882  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

CDS 226...1338  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC17169.1"  
/db\_xref="GI:11343346"  
/translation="MRLIFVYTLICANFCSCRDTSATPOSASIKALRNANLRDEN  
HLTDLVRRDTIQVKGNGYQSPFNPSYPRNLLTWRHLSQENTRIQLVFDNOFGLE  
EAENDICRYDEVEVEDISESTIIRGRCGKHKEVPPRIKSRNQIKITFKSDDYFVAK  
PGKIYVSLLEDFQPAASSETNWSVTSSISGVSYNSPVDPTLIADALDKKIAEFD  
TVEDLLKYPNPSQEDLENNYLDTPRYGRSYHDKRSKVDLDRLNDADKRYSCPTPN  
YSNIREELKLANVFFPRCLLVQRCGNGCGCTVNNRSCNCSGKTVKRYHEVLOFE  
PGHKRRGRAKTALVDIQLDHHERCDCICSSRPR"  
BASE COUNT 566 a 407 c 430 g 479 t  
ORIGIN

Query Match 96.6%; Score 482.8; DB 6; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 9e-80; 2; Indels 4; Gaps 1;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
Qy 1 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
Db 1866 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1807  
Qy 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAAATTTAAAAATTTAAATTTATCTCCTAAATTTT 120  
Db 1806 GCAAGATTCAAAATTTGTTTGTGTTTCAAAAATTTAAAAATTTAAATTTATCTCCTAAATTTT 1747  
Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGTTTAAAGAGTCT 180  
Db 1746 CTAAGACATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGTTTAAAGAGTCT 1687  
Qy 181 AACTCAACATATGTAAGCTCTGCTGACCTGGTGTATATATACCAAAAACATTT 236  
Db 1686 AACTCAACATATGTAAGCTCTGCTGACCTGGTGTATATATACCAAAAACATTT 1627  
Qy 237 TGATCTATATACATAGACATGAATATATTTCTGTGTGTTTGTGTCATATATAACCTC 296  
Db 1626 TGATCTATATACATAGACATGAATATATTTCTGTGTGTTTGTGTCATATATAACCTC 1567  
Qy 297 AAACACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTT 356  
Db 1566 AAACACTATTATTAATGCAATCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTT 1507  
Qy 357 CTACTTGCCATGGCATTAACAAGCAAGCTGAGACTCAGCAACCACTTGTTGTCATTGC 416  
Db 1506 CTACTTGCCATGGCATTAACAAGCAAGCTGAGACTCAGCAACCACTTGTTGTCATTGC 1447  
Qy 417 ATTGCGAGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAAGGTCCTTATCTCACCTCC 476  
Db 1446 ATTGCGAGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAAGGTCCTTATCTCACCTCC 1387  
Qy 477 TTAACACTAAAGGTTCTTTTCAGGCT 500  
Db 1386 TTAACACTAAAGGTTCTTTTCAGGCT 1363

RESULT 15  
AX207483/c  
LOCUS AX207483 1882 bp DNA linear PAT 31-AUG-2001  
DEFINITION Sequence 1 from Patent WO0157083.  
ACCESSION AX207483  
VERSION AX207483.1 GI:15422211  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Gilbertson,D.G. and Hart,C.E.  
TITLE Methods for promoting growth of bone, ligament, and cartilage using  
zvegfr4  
JOURNAL Patent: WO 0157083-A 1 09-AUG-2001;  
ZymoGenetics, Inc. (US)

FEATURES  
source Location/Qualifiers  
1..1882  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
226..1338  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC60334.1"  
/db\_xref="GI:15422212"  
/translation="MRLIFVVTLICANFCSCRDTSATPQASIKALRNANLRDSEN  
HITDLYRRDETIOVKGNGYQSPRPNSYPNLLLTWRLHSENRPIQLVFNQGLE  
EANDICRDFEVEDISETSIIRGWCGHKEVPPIKSRNTQIKITPKSDDIYFAK  
PFKIIYSLLEDQFAAASETNWESVTSSISGVNSPSTVPTLIADALDKKIAEFD  
TVEDLLKYFNPSWQEDLENMYLDTPRYGRSHDRKSKVDLRDLNDADAKRYSCTPRN  
YSVNIREEELKANVFFPRCLLVQRCGNGCGGTWNRSCTCNSGKTVKKYHEVLQFE  
PGHKRRGRKATMALVDIOLDHHERDCICSSRPPR"

BASE COUNT 566 a 407 c 430 g 479 t  
ORIGIN

Query Match 96.6%; Score 482.8; DB 6; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 9e-80;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GTATATGTAAGAAAGCCTCATCTTTTGGATTTTAAATATACAGATGCTTCTTTTAAGAGA 60  
Db |||||||  
QY 61 GCAAGATTCAAAATCTTTTGGTTTCAAAATTTAAATAATATATCTCTAAATTTT 120  
Db |||||||  
QY 1806 GCAAGATTCAAAATCTTTTGGTTTCAAAATTTAAATAATATATCTCTAAATTTT 1747  
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAGAGTCT 180  
Db |||||||  
QY 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAGAGTCT 1687  
QY 181 AACTCAACATATGTAAGCTCTGGTGACCTGCTGTTATATATACCC----AAAAAAACATT 236  
Db |||||||  
QY 1686 AACTCAACATATGTAAGCTCTGGTGACCTGCTGTTATATATACCCAAAAAAACAAATT 1627  
QY 237 TGATCTATATACATAGACATGAATATATTTCTGTGCTGTTTGGCATATATACCTC 296  
Db |||||||  
QY 1626 TGATCTATATACATAGACATGAATATATTTCTGTGCTGTTTGGCATATATACCTC 1567  
QY 297 AAACACTATTTAAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTT 356  
Db |||||||  
QY 1566 AAACACTATTTAAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTT 1507  
QY 357 CTACTTGCATGCGCATTAACAAGCAAGCGCTGAGACTCAGCAACCACTTGTGTTCAATGC 416  
Db |||||||  
QY 1506 CTACTTGCATGCGCATTAACAAGCAAGCGCTGAGACTCAGCAACCACTTGTGTTCAATGC 1447  
QY 417 ATTGCAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAGGGTCTCTTATCTCACCCCTCC 476  
Db |||||||  
QY 1446 ATTGCAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAGGGTCTCTTATCTCACCCCTCC 1387  
QY 477 TTAACCTAAAGGTTCTTTCAGGCT 500  
Db |||||||  
QY 1386 TTAACCTAAAGGTTCTTTCAGGCT 1363

Search completed: July 4, 2003, 16:40:37  
Job time : 1509.66 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.445 Seconds

(without alignments)

6765.003 Million cell updates/sec

US-10-083-853b-2\_COPY\_1\_500

Perfect score: 500

Sequence: 1 gttatgtaagaagcctca.....actaaaggtcttcagggt 500

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first-45-summaries

Database: NCBI Genbank

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:

2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:

3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:

4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:

5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:

6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:

7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:

8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:

9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:

10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:

11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:

12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:

13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:

14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:

15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:

16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:

17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:

18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:

19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:

20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:

21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:

22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:

23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:

24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	500	100.0	1587	22 AAS04493	Human FCTR2 DNA pr
C 2	500	100.0	1728	22 AAS04498	Human PDGFD DNA.
C 3	500	100.0	1828	22 AAS04492	Human FCTR1 DNA pr
C 4	500	100.0	1934	21 AAD00737	Human platelet Der
C 5	500	100.0	2253	21 AAD00738	Human platelet Der
C 6	500	100.0	3718	24 AAD25489	Human LP85 DNA #2.
C 7	500	100.0	3736	24 AAD25488	Human LP85 DNA #1.
C 8	499.6	99.9	4001	22 AAH46957	Human secreted pro
C 9	498.4	99.7	3798	22 AAH46939	Human secreted pro

C 10	497.6	99.5	2726	22 AAH46959	Human secreted pro
C 11	484.8	97.0	3853	22 AAF24196	Human VEGF-G CDNA.
C 12	482.8	96.6	1882	21 AAC81555	Human growth facto
C 13	482.8	96.6	1882	21 AAS51541	SEQ. ID. 36 from W
C 14	482.8	96.6	1882	22 AAH47772	Human zveg4 polyp
C 15	482.8	96.6	1882	22 ABQ73239	Human zveg4 encod
C 16	460	92.0	485	24 ABL63632	Breast cancer rela
C 17	131.8	26.4	360	21 AAD00739	Human platelet Der
C 18	131.8	26.4	620	22 ARA08941	Human novel protei
C 19	109.2	21.8	3121	22 AAF24198	Human VEGF-G CDNA.
C 20	74.2	14.8	1472	21 AAC81596	Mouse growth facto
C 21	74.2	14.8	1472	22 AAH47773	Mouse zveg4 polyp
C 22	74.2	14.8	1472	24 ABQ73240	Mouse zveg4 encod
C 23	62	12.4	690	21 AAD00736	Human platelet Der
C 24	61.4	12.3	256	22 AAS46959	Human breast cance
C 25	61.4	12.3	256	22 AAF17529	Human breast cance
C 26	61.4	12.3	256	24 ABK94995	Human breast tumou
C 27	56.4	11.3	7536	24 ABL33465	Human immune syste
C 28	51.4	10.3	6244	24 ABL32484	Human immune syste
C 29	51.4	10.3	7346	24 ABL32344	Human immune syste
C 30	51.4	10.3	9810	24 ABL32426	Human immune syste
C 31	51	10.2	6167	24 ABK28369	DNA transcription
C 32	49.6	9.9	14551	24 ABL34584	Human metastasis a
C 33	49.4	9.9	19124	18 AAT72882	Plasmodium var-7 p
C 34	49.4	9.9	19124	21 AAZ98287	Human immune syste
C 35	49.2	9.8	5883	24 ABL32352	Human metastasis a
C 36	49.2	9.8	5883	24 ABL34472	Chemically treated
C 37	49.2	9.8	6641	24 ABL54335	Human immune syste
C 38	49.2	9.8	6641	24 ABL32314	Breast cancer rela
C 39	49	9.8	229	24 ABL63535	Breast cancer rela
C 40	49	9.8	229	24 ABL63968	Chemically treated
C 41	49	9.8	6012	24 ABL70328	Human gene regulat
C 42	49	9.8	6012	24 AAS61275	Signal transductio
C 43	49	9.8	6012	24 ABK31371	Human immune syste
C 44	49	9.8	9810	24 ABL32427	Human angiogenesis
C 45	49	9.8	18357	24 ABQ67083	

ALIGNMENTS

RESULT 1

AAS04493/C

ID AAS04493 standard; DNA; 1587 BP.

XX AC AAS04493;

XX DT 07-SEP-2001 (first entry)

XX DE Human FCTR2 DNA present in clone 30664188.0.331.

XX KW Bone morphogenetic protein-1; BMP-1; vascular endothelial growth factor; VEGF-E; platelet derived growth factor; PDGF; FCTR2; hyperplasia; cancer; neoplasia; anaemia; leukopenia; baldness; cardiovascular disorder; fibrotic disorder; diabetic ulcer; obesity; hyperproliferation; human; dysproliferation; neurodegenerative disorder; osteoarthritis; epilepsy; inflammatory disorder; neurodegenerative disorder; coagulation; ds; haemophilia; neural disorder; Parkinson's disease; Alzheimer's disease; multiple sclerosis; Huntington's disease; amyotrophic lateral sclerosis; peripheral neuropathy; acute brain injury.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 540..938

XX FT /\*tag= a

XX FT /product= "Human FCTR2"



XX 07-OCT-1999; 99US-0158083.  
PR 13-OCT-1999; 99US-0159231.  
PR 04-JAN-2000; 2000US-0174485.  
PR 03-MAR-2000; 2000US-0186707.  
PR 10-MAR-2000; 2000US-0188250.  
PR 08-AUG-2000; 2000US-0223879.  
PR 12-SEP-2000; 2000US-0662783.  
PR 20-SEP-2000; 2000US-0234082.  
XX (CURA-) CURAGEN CORP.  
XX Shinkets RA, Lichenstein H, Herrmann JL, Boldog FL, Minskoff S;  
PI Jeffers M;  
XX WPI; 2001-316172/33.  
DR P-PSDB; AAU00699.  
XX Novel growth factor polypeptides termed as FCTR polypeptides, useful  
PT for treating cancer, cardiovascular and fibrotic diseases, diabetic  
PT ulcers, wound healing and neuronal disorders  
XX  
PS Claim 11; Page 14-15; 171pp; English.  
XX The sequence represents DNA encoding a protein related to bone  
CC morphogenetic protein-1 (BMP-1), vascular endothelial growth factor  
CC (VEGF-E) and platelet derived growth factor (PDGF). Polypeptides and  
CC polynucleotides related to BMP-1, VEGF-E and PDGF are referred to as  
CC FCTR polypeptides and nucleic acids. FCTR proteins are useful for treating  
CC or preventing a disorder associated with aberrant expression, aberrant  
CC processing, or aberrant physiological interactions of the proteins in a  
CC mammal, where the disorder is characterised by insufficient or  
CC ineffective growth of a cell or a tissue, e.g. hyperplasia or neoplasia.  
CC The peptides and their associated nucleic acids are useful for both  
CC promoting and inhibiting growth of cells and tissues and in treatment of  
CC cancer, anaemia, leukopenia, baldness, for treating cardiovascular and  
CC fibrotic disorders, diabetic ulcers, obesity, infectious diseases,  
CC hyperproliferative and dysproliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, inflammatory disorders, Graft versus host  
CC disease, coagulation disorders such as haemophilia, and neural disorders  
CC including Parkinson's disease, Alzheimer's disease, multiple sclerosis,  
CC Huntington's disease, amyotrophic lateral sclerosis, peripheral  
CC neuropathy, acute brain injury and epilepsy.  
XX  
SQ Sequence 1587 BP; 522 A; 298 C; 329 G; 438 T; 0 other;  
Query Match 100.0%; Score 500; DB 22; Length 1587;  
Best Local Similarity 100.0%; Pred. No. 3e-96;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTATATGTAGAAAGCCCTCATCTTTTGATTTTAAATACAGATGCTTTCTTTAAGAGA 60  
DB 1462 GTATATGTAGAAAGCCCTCATCTTTTGATTTTAAATACAGATGCTTTCTTTAAGAGA 1403  
QY 61 GCAAGATCAAATTTGTTGTTTCAAAATTTAAATTTATCTCCTAAATTTT 120  
DB 1402 GCAAGATCAAATTTGTTGTTTCAAAATTTAAATTTATCTCCTAAATTTT 1343  
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGTTTAAAGTCT 180  
DB 1342 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGTTTAAAGTCT 1283  
QY 181 AACTCAACATPATGAGCTCTGGTGACCTGGTTATATATACCAAAAAAACATTTGAT 240  
DB 1282 AACTCAACATPATGAGCTCTGGTGACCTGGTTATATATACCAAAAAAACATTTGAT 1223  
QY 241 CTATATACATACATGATATATTTCTGCTGCTGTTTGTGATATATACCTCAAC 300  
DB 1222 CTATATACATACATGATATATTTCTGCTGCTGTTTGTGATATATACCTCAAC 1163  
QY 301 ACTATTATTAAATGCAATCCCTATATCTTAGGTAGAGTTGATGATATACCTTTCTAC 360  
DB 1162 ACTATTATTAAATGCAATCCCTATATCTTAGGTAGAGTTGATGATATACCTTTCTAC 1103

QY 361 TTGCCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTTCATTGCAATTG 420  
DB 1102 TTGCCATGGCATTAACAAGCAAGCTGAGACTCAGCAACACCTTGTTGTTTCATTGCAATTG 1043  
QY 421 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAAGGTCCTTATCTCACCCTCCTTAA 480  
DB 1042 CAGGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAAGGTCCTTATCTCACCCTCCTTAA 983  
QY 481 ACTAAAGGTTCTTTCAGGCT 500  
DB 982 ACTAAAGGTTCTTTCAGGCT 963  
RESULT 2  
AAS04498/C  
ID AAS04498 standard; DNA; 1728 BP.  
XX  
AC AAS04498;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Human PDGFD DNA.  
XX  
KW Bone morphogenetic protein-1; BMP-1; vascular endothelial growth factor;  
KW VEGF-E; platelet derived growth factor; PDGF; FCTR; hyperplasia; cancer;  
KW neoplasia; anaemia; leukopenia; baldness; cardiovascular disorder;  
KW fibrotic disorder; diabetic ulcer; obesity; hyperproliferation; human;  
KW dysproliferation; neurodegenerative disorder; osteoarthritis; epilepsy;  
KW inflammatory disorder; Graft versus host disease; coagulation; ds;  
KW haemophilia; neural disorder; Parkinson's disease; Alzheimer's disease;  
KW multiple sclerosis; Huntington's disease; amyotrophic lateral sclerosis;  
KW peripheral neuropathy; acute brain injury.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 82..1194  
FT FT /\*tag= a  
FT FT /\*product= "Human PDGFD"  
FT FT 1..205  
FT FT /\*tag= b  
FT FT /number= Exon 1  
FT FT 206..410  
FT FT /\*tag= c  
FT FT /number= Exon 2  
FT FT 411..591  
FT FT /\*tag= d  
FT FT /number= Exon 3  
FT FT 592..654  
FT FT /\*tag= e  
FT FT /number= Exon 4  
FT FT 655..853  
FT FT /\*tag= f  
FT FT /number= Exon 5  
FT FT 854..1068  
FT FT /\*tag= g  
FT FT /number= Exon 6  
FT FT 1069..1728  
FT FT /\*tag= h  
FT FT /number= Exon 7  
XX  
PN WO200125437-A2.  
XX  
PD 12-APR-2001.  
XX  
PF 06-OCT-2000; 2000WO-US27671.  
XX  
PR 07-OCT-1999; 99US-0158083.  
PR 13-OCT-1999; 99US-0159231.  
PR 04-JAN-2000; 2000US-0174485.  
PR 03-MAR-2000; 2000US-0186707.  
PR 10-MAR-2000; 2000US-0188250.

PR 08-AUG-2000; 2000US-0223879.  
PR 12-SEP-2000; 2000US-0662783.  
PR 20-SEP-2000; 2000US-0234082.  
PA (CURA-) CURAGEN CORP.  
XX Shimkets RA, Lichenstein H, Herrmann JL, Boldog FL, Minskoff S;  
PI Jeffers M;  
XX WPI: 2001-316172/33.  
DR P-PSDB: AAU00704.  
XX Novel growth factor polypeptides termed as FCTR polypeptides, useful  
PR for treating cancer, cardiovascular and fibrotic diseases, diabetic  
PT ulcers, wound healing and neuronal disorders  
PT  
PS Disclosure; Fig 13; 171pp; English.  
XX  
CC The sequence represents DNA encoding a protein related to bone  
CC morphogenetic protein-1 (BMP-1), vascular endothelial growth factor  
CC (VEGF-E) and platelet derived growth factor (PDGF). Polypeptides and  
CC polynucleotides related to BMP-1, VEGF-E and PDGF are referred to as  
CC FCTR polypeptides and nucleic acids. FCTR proteins are useful for treating  
CC or preventing a disorder associated with aberrant expression, aberrant  
CC processing, or aberrant physiological interactions of the proteins in a  
CC mammal, where the disorder is characterised by insufficient or  
CC ineffective growth of a cell or a tissue, e.g. hyperplasia or neoplasia.  
CC The peptides and their associated nucleic acids are useful for both  
CC promoting and inhibiting growth of cells and tissues and in treatment of  
CC cancer, anaemia, leukopenia, baldness, for treating cardiovascular and  
CC fibrotic disorders, diabetic ulcers, obesity, infectious diseases,  
CC hyperproliferative and dysproliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, inflammatory disorders, graft versus host  
CC disease, coagulation disorders such as haemophilia, and neural disorders  
CC including Parkinson's disease, Alzheimer's disease, multiple sclerosis,  
CC Huntington's disease, amyotrophic lateral sclerosis, peripheral  
CC neuropathy, acute brain injury and epilepsy.  
XX  
SQ Sequence 1728 BP; 530 A; 364 C; 379 G; 455 T; 0 other;  
Query Match 100.0%; Score 500; DB 22; Length 1728;  
Best Local Similarity 100.0%; Pred. No. 3e-96;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAAGATGCTTCTTTAAAGAGA 60  
DB 1718 GTATATCTAAGAAGCCTCATCTTTTCATTTTAAATATACAAAGATGCTTCTTTAAAGAGA 1659  
QY 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCCTAAATTTT 120  
DB 1658 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATTTATCTCCTAAATTTT 1599  
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAAGATTTTAAAGATCT 180  
DB 1598 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAAGATTTTAAAGATCT 1539  
QY 181 AACTCAACATATCTAAGCTCTGCTGTAACCTGTTATATATACAAAACAAATTTGAT 240  
DB 1538 AACTCAACATATGTAAGCTCTGCTGTAACCTGTTATATATACAAAACAAATTTGAT 1479  
QY 241 CTATATACATACATGATGAATATATTTCTGTGTGTTTGTGATATATATACCTCAAC 300  
DB 1478 CTATATACATACATGATGAATATATTTCTGTGTGTTTGTGATATATATACCTCAAC 1419  
QY 301 ACTATTATTAAAGCAATCTTATTTCTAGGTATAGAGTGTGATATACCTTTCTAC 360  
DB 1418 ACTATTATTAAAGCAATCTTATTTCTAGGTATAGAGTGTGATATATACCTTTCTAC 1359  
QY 361 TTGCGATGGCATTAACAAAGCAGGCTGAGACTCAGCAACCTGTGTTTCATTGCAATG 420  
DB 1358 TTGCGATGGCATTAACAAAGCAGGCTGAGACTCAGCAACCTGTGTTTCATTGCAATG 1299  
QY 421 CAGGCTAGTAGTAAGTTTGGTTGCTGTAGGAAAGGGTCTCTTATCTACCCCTCCTTAA 480

DB 1298 CAGGCTAGTAGTAAGTTTGGTTGCTGTAGGAAAGGGTCTCTTATCTACCCCTCCTTAA 1239  
QY 481 ACTAAAGGTTCTTTCAGGCT 500  
DB 1238 ACTAAAGGTTCTTTCAGGCT 1219  
RESULT 3  
AAS04492/C  
ID AAS04492 standard; DNA; 1828 BP.  
XX AAS04492;  
XX AC  
XX AC  
DT 07-SEP-2001 (first entry)  
XX Human FCTR1 DNA present in clone 30564188.0.99.  
XX Bone morphogenetic protein-1; BMP-1; vascular endothelial growth factor;  
KW VEGF-E; platelet derived growth factor; PDGF; FCTR; hyperplasia; cancer;  
KW neoplasia; anaemia; leukopenia; baldness; cardiovascular disorder;  
KW fibrotic disorder; diabetic ulcer; obesity; hyperproliferation; human;  
KW dysproliferation; neurodegenerative disorder; osteoarthritis; epilepsy;  
KW inflammatory disorder; graft versus host disease; coagulation; ds;  
KW haemophilia; neural disorder; Parkinson's disease; Alzheimer's disease;  
KW multiple sclerosis; Huntington's disease; amyotrophic lateral sclerosis;  
KW peripheral neuropathy; acute brain injury.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH 182..1294  
CDS /\*tag= a  
FT /product= "Human FCTR1"  
FT .sig\_peptide 182..250  
FT /\*tag= b  
FT mat\_peptide 251..1291  
FT /\*tag= c  
FT /product= "Mature human FCTR1"  
XX  
XX WO200125437-A2.  
XX  
XX 12-APR-2001.  
XX  
XX 06-OCT-2000; 2000WO-US27671.  
XX  
XX 07-OCT-1999; 99US-0158083.  
PR 13-OCT-1999; 99US-0159231.  
PR 04-JAN-2000; 2000US-0174485.  
PR 03-MAR-2000; 2000US-0186707.  
PR 10-MAR-2000; 2000US-0188250.  
PR 08-AUG-2000; 2000US-0223879.  
PR 12-SEP-2000; 2000US-0662783.  
PR 20-SEP-2000; 2000US-0234082.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Lichenstein H, Herrmann JL, Boldog FL, Minskoff S;  
PI Jeffers M;  
XX WPI: 2001-316172/33.  
DR P-PSDB: AAU00698.  
XX Novel growth factor polypeptides termed as FCTR polypeptides, useful  
PT for treating cancer, cardiovascular and fibrotic diseases, diabetic  
PT ulcers, wound healing and neuronal disorders  
XX  
PS Claim 11; Page 11-12; 171pp; English.  
XX  
CC The sequence represents DNA encoding a protein related to bone  
CC morphogenetic protein-1 (BMP-1), vascular endothelial growth factor  
CC (VEGF-E) and platelet derived growth factor (PDGF). Polypeptides and  
CC polynucleotides related to BMP-1, VEGF-E and PDGF are referred to as

1513 AACACAAATATGTAAGCTCTGGTGTACCTGGTATATATACCATATATAC

; tissue remodelling; treat;

181 AACTCAAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACCAAAAAACATTTGAT 240  
|||||  
1313 AACTCAAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACCAAAAAACATTTGAT 1254

181 AACTCAAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACCAAAAAACATTTGAT 240  
|||||  
1313 AACTCAAACATATGTAAGCTCTGGTGTAACCTGGTTATATATACCAAAAAACATTTGAT 1254









Db 1699 ACTATTATTAATGCAATCTATATTTCTTAGGTATAGAGTTGATGATATACCTTTCTAC 1640  
QY 361 TTGCGATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTTCATTGCAATG 420  
Db 1639 TTGCGATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTTCATTGCAATG 1580  
QY 421 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAGAAAGGGTCTCTATCTCACCTCCTCTAA 480  
Db 1579 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAGAAAGGGTCTCTATCTCACCTCCTCTAA 1520  
QY 481 ACTAAAGGTTCTTTCAGGCT 500  
Db 1519 ACTAAAGGTTCTTTCAGGCT 1500

RESULT 9  
AAH46939/c  
ID AAH46939 standard; cDNA; 3798 BP.  
XX  
AC AAH46939;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human secreted protein encoding cDNA (clone Id HGCNC48).  
XX  
KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neutropic; neuroprotective; antibacterial; virucide; fungicide; human;  
KW ophthalmological; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155430-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01431.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 12-SEP-2000; 2000US-0231968.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;  
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;  
PI Ni J, Ruben SM, Barash SC;  
XX  
DR WPI: 2001-476220/51.  
DR P-PSDB; AAB85529.  
XX  
XX 17 isolated nucleic acid molecules encoding human secreted proteins,  
PT used to preventing, treating or ameliorating a medical condition -  
PT  
XX  
PS Claim 1; Page 417-418; 482pp; English.  
XX

XX The invention provides novel human secreted proteins and polynucleotides  
XX encoding them. The secreted proteins can be expressed by standard  
XX recombinant methodology. The secreted proteins and polynucleotides are  
XX used to prevent, treat or ameliorate a medical condition in e.g. humans,  
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can  
XX also be used in diagnosing a pathological condition. The antibodies to  
XX the proteins can also be used in alleviating symptoms associated with the  
XX disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme  
XX linked immunosorbent assays (ELISA). Disorders which are diagnosed or  
XX treated include autoimmune diseases e.g. rheumatoid arthritis,  
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,  
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
XX e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
XX ocular disorders e.g. corneal infection. The polypeptides can also be  
XX used to aid wound healing and epithelial cell proliferation, to prevent

CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities. The present  
CC sequence represents a human secreted protein encoding cDNA.  
XX  
SQ Sequence 3798 BP; 1189 A; 749 C; 737 G; 1123 T; 0 other;  
Query Match 99.7%; Score 498.4; DB 22; Length 3798;  
Best Local Similarity 99.8%; Pred. No. 7.1e-96;  
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GATATGTGAAGAAAGCCTCATCTTTTGATTTTAAATATATACAAGATGCTTTCTTTAAGAGA 60  
Db 1802 GATATGTGAAGAAAGCCTCATCTTTTGATTTTAAATATATACAAGATGCTTTCTTTAAGAGA 1743  
QY 61 GCAAGATTCAAAATTTGTTTGGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 120  
Db 1742 GCAAGATTCAAAATTTGTTTGGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 1683  
QY 121 CTAAGACATGTTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180  
Db 1682 CTAAGACATGTTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 1623  
QY 181 AACTCAACATATGTAAGCTCTGGTGACCTGTTTATATATATACCAAAAAACATTTGAT 240  
Db 1622 AACTCAACATATGTAAGCTCTGGTGACCTGTTTATATATATACCAAAAAACATTTGAT 1563  
QY 241 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATATACCTCAAC 300  
Db 1562 CTATATACATAGACATGAATATATTTCTGTGTGTTTGGCATATATATACCTCAAC 1503  
QY 301 ACTATTATTAATGCAATCTATATCTTAGGTATAGAAAGTTGATGATATACCTTTCTAC 360  
Db 1502 ACTATTATTAATGCAATCTATATCTTAGGTATAGAAAGTTGATGATATACCTTTCTAC 1443  
QY 361 TTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTTCATTGCAATG 420  
Db 1442 TTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACCACTTGTTGTTTCATTGCAATG 1383  
QY 421 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAGAAAGGGTCTCTATCTCACCTCCTTAA 480  
Db 1382 CAGGCTAGTAGTAAGTTTGGTTGCTGCTAGGAGAAAGGGTCTCTATCTCACCTCCTTAA 1323  
QY 481 ACTAAAGGTTCTTTCAGGCT 500  
Db 1322 ACTAAAGGTTCTTTCAGGCT 1303  
RESULT 10  
AAH46959  
ID AAH46959 standard; cDNA; 2726 BP.  
XX  
AC AAH46959;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human secreted protein encoding cDNA (clone Id HGCNC48).  
XX  
KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neutropic; neuroprotective; antibacterial; virucide; fungicide; human;  
KW ophthalmological; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155430-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01431.  
XX  
PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.  
PR 12-SEP-2000; 2000US-0231968.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;  
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;  
PI Ni J, Ruben SM, Barash SC;  
XX  
XX WPI; 2001-476220/51.  
DR P-PSDB; AAB85549.  
XX  
XX 17 isolated nucleic acid molecules encoding human secreted proteins,  
PT used to preventing, treating or ameliorating a medical condition -  
XX  
XX Claim 1; Page 433; 482pp; English.  
XX  
CC The invention provides novel human secreted proteins and polynucleotides  
CC encoding them. The secreted proteins can be expressed by standard  
CC recombinant methodology. The secreted proteins and polynucleotides are  
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can  
CC also be used in diagnosing a pathological condition. The antibodies to  
CC the proteins can also be used in alleviating symptoms associated with the  
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme  
CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or  
CC treated include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities. The present  
CC sequence represents a human secreted protein encoding cDNA.  
XX  
SQ Sequence 2726 BP; 881 A; 465 C; 489 G; 886 T; 5 other;  
  
Query Match 99.5%; Score 497.6; DB 22; Length 2726;  
Best Local Similarity 99.4%; Pred. NO. 1e-95;  
Matches 497; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAATTTAAATATACAGATGCTTCTTAAGAGA 60  
DB |||||||  
QY 61 GCAAGATTCAAAATTTGTTTCTGTTTCAAAATTTAAATATATACAGATGCTTCTTAAGAGA 120  
DB |||||||  
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 180  
DB |||||||  
QY 2122 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAGAGTCT 2181  
DB |||||||  
QY 181 AACTCAACATATGTAAGCTCTGTTGACCTGGTATATATACCAAAAAACATTTGAT 240  
DB |||||||  
QY 2182 AACTCAACATATGTAAGCTCTGTTGACCTGGTATATATACCAAAAAACATTTGAT 2241  
DB |||||||  
QY 241 CTATATACACATAGACATGAATATATTTCTGTTGTTTGTGATATATACCTCAAC 300  
DB |||||||  
QY 2242 CTATATACACATAGACATGAATATTTCTGTTGTTTGTGATATATACCTCAAC 2301  
DB |||||||  
QY 301 ACTATATTAATGAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360  
DB |||||||  
QY 2302 ACTATATTAATGAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 2361  
DB |||||||  
QY 361 TTGCCATGGCATTAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTTTCATTCATG 420  
DB |||||||  
QY 2362 TTGCCATGGCATTAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTTTCATTCATG 421  
DB |||||||

QY 421 CAGCTAGTAGTAAGTTTGGTTGCTGTAGCAAAAGGTCCTCTATCTACCCCTCTTAA 480  
DB |||||||  
QY 2422 CAGCTAGTAGTAAGTTTGGTTGCTGTAGCAAAAGGTCCTCTATCTACCCCTCTTAA 2481  
DB |||||||  
QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
DB |||||||  
QY 2482 ACTAATGGTTCTTTTCAGGCT 2501  
DB |||||||  
ID AAF24196 standard; DNA; 3853 BP.  
XX AC AAF24196;  
XX  
DT 02-APR-2001 (first entry)  
DB Human VEGF-G cDNA.  
XX KW Vascular endothelial growth factor; VEGF; cancer; cell;  
KW angiogenesis; ss.  
XX OS Homo sapiens.  
XX PN W020000087/8-A2  
XX DP 01 JAN-2001.  
XX PF 29-JUN-2000; 2000WO-US18085.  
XX PR 30-JUN-1999; 99US-0343671.  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX PI Gearing DP;  
XX DR WPI; 2001-050129/06.  
XX  
PT New vascular endothelial growth factor family member used for diagnosis  
PT and treatment of deregulated cell growth e.g. cancer, disorders  
PT involving aberrant angiogenesis e.g. psoriasis, and chronic  
PT inflammatory diseases -  
XX  
PS Claim 1; Fig 1; 142pp; English.  
XX  
CC The present invention relates to a vascular endothelial growth factor  
CC (VEGF) family member, VEGF-G. VEGF-G protein and nucleic acid  
CC molecules are used as modulating agents or as targets for  
CC developing modulating agents to regulate a variety of cellular  
CC processes e.g. cell proliferation, differentiation, migration and  
CC wound repair. VEGF-G modulators, i.e. VEGF-G protein, peptide,  
CC peptidomimetic or nucleic acid are used to treat a subject with  
CC aberrant VEGF-G protein or nucleic acid expression or activity  
CC e.g. deregulated cell growth, such as cancer, hyperproliferic bone  
CC disorders, disorders involving aberrant angiogenesis e.g. psoriasis,  
CC and chronic inflammatory diseases e.g. rheumatoid arthritis. VEGF-G  
CC gene expression is inhibited through the administration of antisense  
CC molecules or ribozymes and by targeting the regulatory region of VEGF-G  
CC to prevent transcription of the gene in target cells.  
SQ Sequence 3853 BP; 1194 A; 771 C; 757 G; 1131 T; 0 other;  
  
Query Match 99.5%; Score 484.8; DB 22; Length 3853;  
Best Local Similarity 99.4%; Pred. NO. 5.2e-93;  
Matches 498; Conservative 10; Mismatches 2; Indels 2; Gaps 1;  
  
QY 1 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAATTTAAATATACAGATGCTTCTTAAGAGA 60  
DB |||||||  
QY 1851 GTATATGTAAGAAAGCCCTCATCTTTTGGTTTCAAAATTTAAATATACAGATGCTTCTTAAGAGA 1792  
DB |||||||  
QY 61 GCAAGATTCAAAATTTGTTTCTGTTTCAAAATTTAAATATATCTCCTCTTAAATTT 120  
DB |||||||  
QY 1791 GCAAGATTCAAAATTTGTTTCTGTTTCAAAATTTAAATATATCTCCTCTTAAATTT 1732  
DB |||||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180  
DB 1731 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1672  
QY 181 AACTCAACATATGTAAGCTCTGGTGACCTGGTGTATATATAC--AAAAAACATTTG 238  
DB 1671 AACTCAACATATGTAAGCTCTGGTGACCTGGTGTATATATACCAAAAAAACATTTG 1612  
QY 239 ATCTATATACATAGACATGAATATTTCTGTGCTGTTTGGCATATATACCTCAA 298  
DB 1611 ATCTATATACATAGACATGAATATTTCTGTGCTGTTTGGCATATATACCTCAA 1552  
QY 299 ACCTATTATTAATGCAATCCCTATATTTAGGTATAGAGTGTATATACCTTTCT 358  
DB 1551 ACCTATTATTAATGCAATCCCTATATTTAGGTATAGAGTGTATATACCTTTCT 1492  
QY 359 ACTTGCCATGCGATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTTGTTTATGTCAT 418  
DB 1491 ACTTGCCATGCGATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTTGTTTATGTCAT 1432  
QY 419 TGCAGGCTAGTGTAGTGTGTTGTTGCTGTAGGAAAGGCTCTTATCTCACCCCTCTT 478  
DB 1431 TGCAGGCTAGTGTAGTGTGTTGTTGCTGTAGGAAAGGCTCTTATCTCACCCCTCTT 1372  
QY 479 AAACATAAGGTTCTTTCAGGCT 500  
DB 1371 AAACATAAGGTTCTTTCAGGCT 1350

## RESULT 12

AAC81555/c

ID AAC81555 standard; cDNA; 1882 BP.

XX AAC81555;

XX 09-MAR-2001 (first entry)

DE Human growth factor homologue zveg4 cDNA, SEQ ID NO:1.

KW Human; zveg4; growth factor homologue; VEGF/PDGF family;  
KW CUB domain; PDGF-like activity; mitogenic; osteogenic;  
KW neovascularisation; tissue repair; proliferation; differentiation;  
KW liver damage; neurodegenerative; Alzheimer's disease; multiple sclerosis;  
KW periodontal disease; bone fracture; wound healing; vulnerability; ischaemia;  
KW immunomodulation; hepatic; chromosome 11q22.3-23.1; ss.

XX Homo sapiens.

OS WO20006736-A1.

XX 09-NOV-2000.

XX 03-MAY-2000; 2000WO-US40047.

XX 03-MAY-1999; 99US-0304216.

XX 10-NOV-1999; 99US-0164463.

XX 04-FEB-2000; 2000US-0180169.

XX (ZYMO ) ZYMOGENETICS INC.

XX Gilbert T, Hart CE, Sheppard PO, Gilbertson DG;

XX WPI; 2000-687541/67.

XX P-PSDB; AAB48653.

XX Growth factor homologs and the nucleic acids that encode them, useful  
XX e.g. for treating liver damage, ischemia, multiple sclerosis and  
XX Alzheimer's disease -

XX Claim 35; Page 106-110; 143pp; English.

XX The invention relates to the human growth factor homologue zveg4

CC (AAB48653), and nucleic acids encoding it (AAC81555). Zveg4 is a member  
CC of the PDGF (platelet-derived growth factor)/VEGF (vascular endothelial  
CC growth factor) family. Zveg4 has a growth factor domain (AAB48654)  
CC characterised by a PDGF cysteine knot structure, and a CUB domain  
CC (AAB48655) which has a beta barrel structure. Zveg4 has PDGF-like  
CC activity, having mitogenic activity on fibroblasts, vascular smooth  
CC muscle cells and pericytes, and has also been shown to stimulate bone  
CC growth. The invention also relates to fusion proteins comprising human  
CC zveg4 or fragments thereof, particularly human zveg4/human zveg3  
CC fusions; expression constructs and host cells comprising human zveg4  
CC nucleic acids; the recombinant expression of human zveg4; an antibody  
CC which binds to human zveg4 or a fragment thereof; a method of activating  
CC a cell-surface PDGF receptor using a zveg4-derived polypeptide; a  
CC method of modulating the proliferation, differentiation, migration or  
CC metabolism of bone cells, comprising exposing bone cells to  
CC zveg4-derived polypeptides; and a method of detecting a genetic  
CC abnormality in the zveg4 gene of a patient. Zveg4 proteins and derived  
CC fragments may be used to stimulate tissue development or repair, or  
CC cellular differentiation or proliferation. They are particularly used for  
CC the treatment or repair of liver damage, and may also be used to  
CC modulate neurite growth (e.g., in the treatment of Alzheimer's disease or  
CC multiple sclerosis). Due to their osteogenic activity, they may be used  
CC in the treatment of periodontal disease and fractures. They may also be  
CC used to enhance expansion and mobilisation of haematopoietic stem cells  
CC and endothelial precursor stem cells, which may be useful in the  
CC treatment of ischaemia, in wound healing, and in the modulation of the  
CC immune system. The present sequence represents cDNA encoding human  
CC zveg4.  
XX

SQ Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;

Query Match 96.6%; Score 482.8; DB 21; Length 1882;

Best Local Similarity 98.8%; Pred. No. 1.3e-92;

Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GTATATGTAAAGAGCTCATCTTTTGTGTTTCAAAATTTAAATAAATTTATCTCTTAAAGAGA 60  
DB 1866 GTATATGTAAAGAGCTCATCTTTTGTGTTTCAAAATTTAAATAAATTTATCTCTTAAAGAGA 1807  
QY 61 GCAAGATTCAAAATTTGTTGTTTCAAAATTTAAATAAATTTATCTCTTAAATTTT 120  
DB 1806 GCAAGATTCAAAATTTGTTGTTTCAAAATTTAAATAAATTTATCTCTTAAATTTT 1747  
QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180  
DB 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 1687  
QY 181 AACTCAACATATGTAAGCTCTGGTCTGCTGTTATATATATACC---AAAAAACATTT 236  
DB 1686 AACTCAACATATGTAAGCTCTGGTCTGCTGTTATATATATATATATATATATATAT 1627  
QY 237 TGATCTATATACATAGACATGAATATATTTCTGTGTTGTTGTCATATATAACCTC 296  
DB 1626 TGATCTATATACATAGACATGAATATATTTCTGTGTTGTTGTCATATATAACCTC 1567  
QY 297 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTGTGATATACCTTT 356  
DB 1566 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTGTGATATACCTTT 1507  
QY 357 CTACTTGGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTGTGTTTCATTGC 416  
DB 1506 CTACTTGGCCATGCAATTAACAAGCAAGGCTGAGACTCAGCAACCACTGTGTTTCATTGC 1447  
QY 417 ATTGCAGGCTAGTAGTAAGTGTGCTGGTAGGAAAGGGTCTCTTATCTCACCTCC 476  
DB 1446 ATTGCAGGCTAGTAGTAAGTGTGCTGGTAGGAAAGGGTCTCTTATCTCACCTCC 1387  
QY 477 TTAACATAAAGGTTCTTTTCAGGCT 500  
DB 1386 TTAACATAAAGGTTCTTTTCAGGCT 1363

RESULT 13



AAAS1541/c	
ID	AAA51541 standard; cDNA; 1882 BP.
XX	
XX	AAA51541.
XX	
DT	26-SEP-2000 (first entry)
XX	
DE	SEQ. ID. 36 from W00034474.
XX	
KW	Vascular endothelial growth factor; homologue; zveg3; CUB domain;
KW	Cysteine knot; platelet-derived growth factor; PDGF; neuropilin;
KW	chromosome 4q28.3; cystostatic; anti-psoriatic; anti-inflammatory;
KW	anti-diabetic; ophthalmological; anti-rheumatic; anti-arthritis;
KW	vulnerary; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200034474-A2.
XX	
PH	Key
FT	CDS
FT	226..1338
FT	/tag= a
XX	
XX	15-JUN-2000.
XX	
XX	07-DEC-1999; 99WO-US28968.
XX	
PR	07-DEC-1998; 98US-0207120.
PR	06-JUL-1999; 99US-0142576.
PR	21-OCT-1999; 99US-0161653.
PR	12-NOV-1999; 99US-0165255.
XX	
PA	(ZYMO ) ZYMOGENETICS INC.
XX	
PI	Gao Z, Hart CB, Piddington CS, Sheppard PO, Shoemaker KE;
PI	Gilbertson DG, West JW;
XX	
DR	WPI; 2000-423420/36.
DR	P-PSDB; AA956864.
XX	
PT	Novel zveg3 polypeptides and nucleotides encoding them useful for
PT	stimulating growth of smooth muscle cells and fibroblasts comprising an
PT	epitope bearing portion of a specific amino acid sequence
XX	
PS	Disclosure; Page 161-164; 173pp; English.
XX	
CC	Polypeptides comprising an epitope-bearing portion human or murine
CC	zVEGF3 (vascular endothelial growth factor homologue) are claimed. The
CC	growth factors comprise a growth factor domain and a CUB domain (generic
CC	sequence motifs are shown in AA956859 and AA956860). The growth factor
CC	domain is characterized by an arrangement of cysteine residues and
CC	beta-strands that is characteristic of the "cysteine knot" structure of
CC	the platelet-derived growth factor (PDGF) family. The CUB domain shows
CC	homology to CUB domains in neuropilins, human bone morphogenetic
CC	protein-1, porcine seminal plasma protein, bovine acidic seminal fluid
CC	protein and Xenopus laevis tollid-like protein. Structural analysis and
CC	homology predict that zVEGF3 polypeptides complex with a second
CC	polypeptide to form multimeric proteins. The human zveg3 gene has been
CC	mapped to chromosome 4q28.3. zVEGF3 is useful for stimulating the growth
CC	of fibroblasts or smooth muscles cells, for activating cell surface
CC	PDGF-alpha receptor and for inhibiting PDGF-alpha receptor mediated
CC	cellular processes. zVEGF3 is useful for regulating (post-development)
CC	organ growth, regeneration and maintenance, as well as tissue
CC	maintenance and repair processes. zVEGF3 antagonists are useful for
CC	treating cancer, rheumatoid arthritis, diabetic retinopathy, ischemic
CC	limb disease, peripheral vascular disease, myocardial ischemia, vascular
CC	intimal hyperplasia, atherosclerosis, wound healing, chronic liver
CC	disease and haemangioma formation. zVEGF3 can also be used to modulate
CC	neurite growth and development of the nervous system, and for treating
CC	neurodegenerative diseases.
XX	
SQ	Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other:



PR 04-FEB-2000; 2000US-180169P.  
XX 31-MAR-2000; 2000US-054022A.  
PA (ZYMO ) ZYMOGENETICS INC.  
PI Gilbertson DG, Hart CE;  
XX  
XX WPI; 2001-611088/70.  
DR P-PSDB; AAG65601.  
XX  
PT Use of zveg4 polypeptide for promoting bone, ligament or cartilage  
PT growth in mammal at site of fracture, implant, and bone graft, and for  
PT promoting growth or differentiation of osteoblasts, chondrocytes in  
PT culture  
XX  
PS Example 1; Page 44-47; 57pp; English.  
XX  
CC The invention relates to the use of zveg4 polypeptide for promoting  
CC bone, ligament or cartilage growth in a mammal, and for promoting  
CC proliferation or differentiation of osteoblasts, osteoclasts,  
CC chondrocytes or bone marrow stem cells in culture. For promoting  
CC cartilage growth, chondrocytes are cultured ex vivo in presence of the  
CC zveg4 polypeptide and then placed into mammal where cartilage is to be  
CC grown. Zveg4 polypeptide is useful for promoting growth of bone.  
CC ligament or cartilage in a mammal at a site of bony defect such as  
CC fracture, bone graft, implant or periodontal pocket, in humans and non-  
CC human animals such as domestic animals including livestock and companion  
CC animals. Zveg4 is used for promoting growth of bone, ligament, or  
CC cartilage in conditions of bone defects following therapeutic treatments  
CC of bone cancers or other conditions characterized by increased bone loss,  
CC or decreased bone formation, or elevation of peak bone mass in pre-  
CC menopausal woman. It is also useful for healing bone following radiation  
CC -induced osteonecrosis, repairing bone defects arising from surgery, and  
CC promotion of bone healing in plastic surgery, increasing bone formation  
CC during distraction osteogenesis, treating bone injuries including repair  
CC of cartilage and ligament and treatment of osteoporosis. The present  
CC sequence represents a human zveg4 polypeptide encoding cDNA.  
XX  
XX Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;  
XX  
Query Match 96.6%; Score 482.8; DB 22; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 1.3e-92;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
QY 1 GTATATGTAAGAAGCCTCACTTTTGGTTTAAATACAAAGATGCTTTCTTTAGAGA 60  
DB 1866 GTATATGTAAGAAGCCTCACTTTTGGTTTAAATACAAAGATGCTTTCTTTAGAGA 1807  
QY 61 GCAGATTCAAATTTGTTTGTCTTCAAAATTTAAATAATTTATCTCCTAAATTTT 120  
DB 1806 GCAAGATTCAAATTTGTTTGTCTTCAAAATTTAAATAATTTATCTCCTAAATTTT 1747  
QY 121 CTAAGACATGTTTCATATATTTGACATCCCTTATTTTGGCAAAGATTTTAAAGAGTCT 180  
DB 1746 CTAAGACATGTTTCATATATTTGACATCCCTTATTTTGGCAAAGATTTTAAAGAGTCT 1687  
QY 181 AACTCAACATATGTAAAGCTCTGGTGACCTGGTTATATATACCC----AAAAAACAATT 236  
DB 1686 AACTCAACATATGTAAAGCTCTGGTGACCTGGTTATATATACCCAAAAAACAATT 1627  
QY 237 TGATCATATACACATAGACATGAATATATTTCTGTGTGTGTGTGCATATATACCTC 296  
DB 1626 TGATCATATACACATAGACATGAATATATTTCTGTGTGTGTGTGCATATATACCTC 1567  
QY 297 AAACACTATTATTAATGCAATCCTATATCTTAGGTATAGAGTGTGATCATATACCTTT 356  
DB 1566 AAACACTATTATTAATGCAATCCTATATCTTAGGTATAGAGTGTGATCATATACCTTT 1507  
QY 357 CTACTTGCCATGCAATTAACAAAGCCTGAGACTCAGCAACCACTTGTTGTTATTCG 416  
DB 1506 CTACTTGCCATGCAATTAACAAAGCCTGAGACTCAGCAACCACTTGTTGTTATTCG 1447  
QY 417 ATTGCAGGCTAGTAGTAAGTTTGGTTCGTGTAGGAAAGGGTCTCTTATCTCACCCCTC 476

Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTTCGTGTAGGAAAGGGTCTCTTATCTCACCCCTC 1387  
QY 477 TTAAACTAAAGGTTCTTTTCAGGCT 500  
DB 1386 TTAAACTAAAGGTTCTTTTCAGGCT 1363  
RESULT 15  
ABQ73239/C  
ID ABQ73239 standard; cDNA; 1882 BP.  
XX  
XX AC ABQ73239;  
XX  
XX 30-SEP-2002 (first entry)  
XX  
XX Human zveg4 encoding cDNA SEQ ID NO:1.  
XX  
XX Human; zveg4; cell proliferation; extracellular matrix production;  
KW fibroproliferative disorder; PDGF-D; platelet derived growth factor;  
KW PDGF; vascular endothelial growth factor; VEGF; cytotactic; nephrotropic;  
KW hepatotropic; antiinflammatory; osteopathic; antiarthritic; metatarsalgia;  
KW prostate tumour; prostate cancer; glomerulonephritis; lupus nephritis;  
KW diabetic glomerulosclerosis; renal arteriosclerosis; nephrotic syndrome;  
KW chronic active hepatitis; cirrhosis; osteopetrosis; osteosclerosis;  
KW hyperostosis; osteoarthritis; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
CDS 226..1338  
FT /\*tag= a  
FT /product= "zveg4"  
XX  
XX US2002064832-A1.  
XX  
XX 30-MAY-2002.  
XX  
XX 14-MAR-2001; 2001US-0808972.  
XX  
XX 03-MAY-1999; 99US-132250P.  
PR 10-NOV-1999; 99US-164463P.  
PR 04-FEB-2000; 2000US-180169P.  
PR 26-SEP-2000; 2000US-235295P.  
PR 03-MAY-2000; 2000US-0564595.  
XX  
XX (HART/) HART C E.  
PA (TOPO/) TOPOUZIS S.  
PA (GILB/) GILBERTSON D G.  
XX  
XX Hart CE, Topouzis S, Gilbertson DG;  
XX  
XX WPI; 2002-573696/61.  
DR P-PSDB; ABP51640.  
XX  
XX Reducing proliferation or extracellular matrix production by a cell in  
PT a mammal, useful for treating fibroproliferative disorders of bone,  
PT liver and kidney, comprises administering a zveg4 antagonist  
XX  
XX Example 1; Page 18-19; 34pp; English.  
XX  
XX The present invention describes a method for reducing proliferation of  
CC or extracellular matrix production by a cell in a mammal. The method  
CC comprises administering to the mammal a composition comprising a  
CC therapeutically effective amount of a zveg4 antagonist chosen from  
CC anti-zveg4 antibodies, inhibitory polynucleotides, inhibitors of  
CC zveg4 activation, and mitogenically inactive, receptor-binding variants  
CC of zveg4. Zveg4 (also called PDGF-D) is a multi-domain protein that is  
CC structurally related to platelet derived growth factor (PDGF) and  
CC vascular endothelial growth factors (VEGF). Zveg4 has cytostatic,  
CC nephrotropic, hepatotropic, antiinflammatory, osteopathic and  
CC antiarthritic activities. The method is useful for reducing proliferation  
CC of mesangial, epithelial, endothelial, smooth muscle, fibroblast,

CC osteoblast, osteoclast, neuronal, stromal, stellate or interstitial cells  
CC in a mammal, in particular proliferation of prostate tumour cells, and  
CC for reducing extracellular matrix production by a cell in a mammal  
CC suffering from a fibroproliferative disorder of kidney, bone or liver.  
CC In particular it is useful for reducing stellate cell activation. The  
CC method is useful for reducing metastasis of prostate cancer cells to  
CC bone in a mammal and for treating a fibroproliferative disorder of the  
CC kidney, liver or bone in a mammal. Fibroproliferative disorders of the  
CC kidney include, glomerulonephritis, diabetic glomerulosclerosis, lupus  
CC nephritis, renal arteriosclerosis, and nephrotic syndrome, disorders of  
CC the liver include chronic active hepatitis and many other types of  
CC cirrhosis, and disorders of the bone include osteopetrosis, hyperostosis,  
CC osteosclerosis, osteoarthritis, and ectopic bone formation in metastatic  
CC prostate cancer. The present sequence encodes human zveg14, which is  
CC used in an example from the present invention.

XX

SQ Sequence 1882 BP; 566 A; 407 C; 430 G; 479 T; 0 other;

Query Match 96.6%; Score 482.8; DB 24; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 1.3e-92;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY	1	GTATATGTAAGAAAGCCCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA	60
DB	1866	GTATATGTAAGAAAGCCCTCATCTTTTGATTTTATATACAGATGCTTCTTTAAGAGA	1807
QY	61	GCAAGATTCAAATTTGTTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT	120
DB	1806	GCAAGATTCAAATTTGTTGTTTCAAAATTTAAATTAATTTATCTCTAAATTTT	1747
QY	121	CTAAGAGATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT	180
DB	1746	CTAAGAGATGTTTCATATATTGACCATCCCTTATTTGGCAAGAGATTTTAAGAGTCT	1687
QY	181	AACTCAACATATGTAAGCTCTGGTGACCTGGTTATATATAC	236
DB	1686	AACTCAACATATGTAAGCTCTGGTGACCTGGTTATATATAC	1627
QY	237	TGATCTATATACATAGACATGAATATATTCTGTGTGTGTGTGCATATATAACCTC	296
DB	1626	TGATCTATATACATAGACATGAATATATTCTGTGTGTGTGTGCATATATAACCTC	1567
QY	297	AAACACTATTATAATGCAATCCCTATATTCTAGGTATAGAAGTTGATGATATACCTTT	356
DB	1566	AAACACTATTATAATGCAATCCCTATATTCTAGGTATAGAAGTTGATGATATACCTTT	1507
QY	357	CTACTTGGCATGGCAATTAACAAAGCAAGCTGAGACTCAGCAACCACTTGTTTCATTGC	416
DB	1506	CTACTTGGCATGGCAATTAACAAAGCAAGCTGAGACTCAGCAACCACTTGTTTCATTGC	1447
QY	417	ATTGCAAGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGCTCTCTTATCTCACCCCTCC	476
DB	1446	ATTGCAAGCTAGTAGTAAGTTGGTTGCTGGTAGGAAAGGCTCTCTTATCTCACCCCTCC	1387
QY	477	TTAAACTAAAGGTTCTTTCAAGCT	500
DB	1386	TTAAACTAAAGGTTCTTTCAAGCT	1363

Search completed: July 4, 2003, 15:25:01  
Job time : 168.445 secs



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	482.8	96.6	1882	4	US-09-540-224-1	Sequence 1, Appli	
C 2	482.8	96.6	1910	4	US-09-457-066-36	Sequence 36, Appli	
C 3	74.2	14.8	1472	4	US-09-540-224-3	Sequence 3, Appli	
C 4	61.4	12.3	256	4	US-09-222-575-13	Sequence 13, Appli	
C 5	49.4	9.9	19124	2	US-08-487-826B-13	Sequence 13, Appli	
C 6	42	8.4	19124	2	US-08-487-826B-13	Sequence 13, Appli	
C 7	40.8	8.2	1422	1	US-08-319-704-5	Sequence 5, Appli	
C 8	40.2	8.0	162450	4	US-09-345-883-2	Sequence 1, Appli	
C 9	40	8.0	2435	4	US-09-306-593-1	Sequence 1, Appli	
C 10	39.8	8.0	168575	4	US-09-426-290-1	Sequence 1, Appli	
C 11	39.6	7.9	401	4	US-09-643-597-262	Sequence 262, App	
C 12	39.6	7.9	5433	3	US-08-929-329-1	Sequence 1, Appli	
C 13	39.4	7.9	1368	3	US-08-874-563-5	Sequence 5, Appli	
C 14	39.4	7.9	1368	3	US-08-577-483-14	Sequence 14, Appli	
C 15	39.2	7.8	6152	4	US-08-973-462-1	Sequence 1, Appli	
C 16	39	7.8	6350	4	US-09-647-826-1	Sequence 1, Appli	
C 17	38.8	7.8	162450	4	US-09-345-882-1	Sequence 1, Appli	
C 18	38.6	7.7	1016	1	US-08-109-391A-3	Sequence 3, Appli	
C 19	38.6	7.7	1016	1	US-08-459-019A-3	Sequence 3, Appli	
C 20	38.6	7.7	1016	2	US-08-460-428A-3	Sequence 3, Appli	
C 21	38.6	7.7	1016	3	US-08-458-860A-3	Sequence 3, Appli	
C 22	38.6	7.7	2659	3	US-08-749-522-1	Sequence 1, Appli	
C 23	38.6	7.7	2755	3	US-08-749-523-2	Sequence 2, Appli	
C 24	38.6	7.7	8920	2	US-08-446-855A-1	Sequence 1, Appli	
C 25	38.6	7.7	8920	2	US-08-446-855A-1	Sequence 1, Appli	
C 26	38.6	7.7	8920	4	US-09-150-741-1	Sequence 1, Appli	
C 27	38.6	7.7	8920	4	US-09-150-741-1	Sequence 1, Appli	

Db 1566 AAACACTATTATTAAATGCAATCCTATATTCTTAGGTATAGAAGTTGATGATATACCTTT 1507  
Qy 357 CTACTTGCCATGCCATTAAACAAAGCAGGCTGAGACTCAGCAACCACTTCTGTTTCATTGC 416  
Db 1506 CTACTTGCCATGCCATTAAACAAAGCAGGCTGAGACTCAGCAACCACTTCTGTTTCATTGC 1447  
Qy 417 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAGGCTCTTATCTCACCTCC 476  
Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAGGCTCTTATCTCACCTCC 1387  
Qy 477 TTAACATAAGGTTCTTTCAGGCT 500  
Db 1386 TTAACATAAGGTTCTTTCAGGCT 1363

#### RESULT 2

US-09-457-066-36/c  
; Sequence 36, Application US/09457066  
; Patent No. 6432673  
; GENERAL INFORMATION:  
; APPLICANT: Gao, Zeren  
; APPLICANT: Piddington, Christopher S.  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Shoemaker, Kimberly E.  
; APPLICANT: Gilbertson, Debra G.  
; APPLICANT: West, James W.  
; TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3  
; FILE REFERENCE: 98-60  
; CURRENT APPLICATION NUMBER: US/09/457,066  
; CURRENT FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36  
; LENGTH: 1882  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (236)...(1338)  
US-09-457-066-36

Query Match 96.6%; Score 482.8; DB 4; Length 1910;  
Best Local Similarity 98.8%; Pred. No. 8.7e-107;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1 GTATATCTAAGAAGCCTCATCTTTTCATTTTTAAATATACAAGATGCTTTCTTAAGAGA 60  
Db 1894 GTATATCTAAGAAGCCTCATCTTTTCATTTTTAAATATACAAGATGCTTTCTTAAGAGA 1835  
Qy 61 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATATATAATTTATCTCCTTAAATTTT 120  
Db 1834 GCAAGATTCAAAATTTGTTTGTGTTTCAAAATTTAAATATATAATTTATCTCCTTAAATTTT 1775  
Qy 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGATTTTAAAGTCT 180  
Db 1774 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGATTTTAAAGTCT 1715  
Qy 181 AACTCAACATATGTAGCTCTGCTGCTACCTGCTATATATACC---AAAAAACATT 236  
Db 1714 AACTCAACATATGTAGCTCTGCTGCTACCTGCTATATATACC---AAAAAACATT 1655  
Qy 237 TGATCTATATACATAGACATGAATATATTTCTGTTGTTGTTGTCATATATAACCTC 296  
Db 1654 TGATCTATATACATAGACATGAATATATTTCTGTTGTTGTTGTCATATATAACCTC 1595  
Qy 297 AAACACTATTATTAAATGCAATCCTATATTCTTAGGTATAGAAGTTGATGATATACCTTT 356  
Db 1594 AAACACTATTATTAAATGCAATCCTATATTCTTAGGTATAGAAGTTGATGATATACCTTT 1535  
Qy 357 CTACTTGCCATGCCATTAAACAAAGCAGGCTGAGACTCAGCAACCACTTCTGTTTCATTGC 416

Db 1534 CTACTTGCCATGCCATTAAACAAAGCAGGCTGAGACTCAGCAACCACTTCTGTTTCATTGC 1475  
Qy 417 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAGGCTCTTATCTCACCTCC 476  
Db 1474 ATTGCAGGCTAGTAGTAAGTTTGGTTGCTGGTAGGAAAGGCTCTTATCTCACCTCC 1415  
Qy 477 TTAACATAAGGTTCTTTCAGGCT 500  
Db 1414 TTAACATAAGGTTCTTTCAGGCT 1391

#### RESULT 3

US-09-540-224-3/c  
; Sequence 3, Application US/09540224  
; Patent No. 6468543  
; GENERAL INFORMATION:  
; APPLICANT: Gilbertson, Debra G.  
; APPLICANT: Hart, Charles E.  
; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,  
; FILE REFERENCE: 00-28  
; CURRENT APPLICATION NUMBER: US/09/540,224  
; CURRENT FILING DATE: 2000-03-31  
; EARLIER APPLICATION NUMBER: US 60/180,169  
; EARLIER FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1472  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (93)...(1205)  
US-09-540-224-3

Query Match 14.8%; Score 74.2; DB 4; Length 1472;  
Best Local Similarity 69.2%; Pred. No. 2.7e-09;  
Matches 117; Conservative 0; Mismatches 48; Indels 4; Gaps 1;

Qy 215 TATATATACCAAAAAAATTTGATCTATATACATAGACATGAATATTTCTGTGT 274  
Db 1469 TAAATATCTGAATAAGTATTAACTATTAC---ATACATAGAGCTAGTCATCTAT 1414  
Qy 275 GTGTTTGTGCATATATAACCTCAACACTATTATTAAATGCAATCTTATTTCTTAGGTA 334  
Db 1413 ATAAGTATATACATATCTTCACATACTATTGATAAGTGAATTCCTTATGTTATGTATA 1354  
Qy 335 TAGAAGTTGATGATATACCTTTCTACTTGCCATGGCATTAAACAAGCAA 383  
Db 1353 TAGAATTAATGATATTTGCTTCTTGCCATGGCATTAAACAAGCAA 1305

#### RESULT 4

US-09-222-575-13  
; Sequence 13, Application US/09222575  
; Patent No. 6387697  
; GENERAL INFORMATION:  
; APPLICANT: Yugu, Jiang  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer  
; FILE REFERENCE: 210121.470  
; CURRENT APPLICATION NUMBER: US/09/222,575  
; CURRENT FILING DATE: 1998-12-28  
; NUMBER OF SEQ ID NOS: 174  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 256  
; TYPE: DNA  
; ORGANISM: Human







```
NAME/KEY: allele
LOCATION: 108106 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817 : polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817 : polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096 : polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096 : polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865 : complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865 : complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736 : polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736 : polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145 : polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145 : polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177 : polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177 : polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121 : polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121 : polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140 : polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140 : polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828 : polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828 : polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966 : polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966 : polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130 : polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130 : polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177 : polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177 : polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
```

Query Match 8.0%; Score 40.2; DB 4; Length 162450;  
Best Local Similarity 56.4%; Pred. No. 0.97; Mismatches 0; Gaps 0;  
Matches 75; Conservative 0



```
;
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(401)
; OTHER INFORMATION: n = A,T,C or G
US-09-643-597-262

Query Match
Best Local Similarity 7.9%; Score 39.6; DB 4; Length 401;
Matches 144; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 63 AAGATTCAAATTTGTTGTGTTTCAAAATTTAAATAATTTATCTCTCTAAATTTCT 122
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8 AACATTTCTAATATTTTGGCTTTTCATATATCAAGGAGATTATGTGAACATATTTTAA 67
QY 123 AAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAAGATTTTAGAGCTCAA 182
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 ATACTGTAAGTGACATATAGTTATAAGATATATTTCTGTACAGTAGAGAAAGATTAT 127
QY 183 CTCAAACATATGTAAGCTCTGTGTACCTGGTTATATATACCAAAAAAACATTTGATCT 242
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 ACATCAACAATATTTGTACCATATACATTTTCATCTCGATCTCATAGAAATTCAAAA 187
QY 243 ATATACATACATGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTCAACAC 302
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 GAATAATGATACAGGTGAAATATGTTTACTTTCTTAAATCAAGCCTAGTTGTCACTC 247
QY 303 TATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATATACCTTTCTACTT 362
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 AAAAATATGNTGCACTAGTTTATTTTGAATTTAGGTTTGGGACTACTTTTTCCTCANT 307
QY 363 GCATGGCATTAACAAAGCAA 383
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 TCAATGAGAAATAAAATCTA 328

RESULT 12
US-08-929-329-1
; Sequence 1, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 1:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
US-08-929-329-1

Query Match
Best Local Similarity 7.9%; Score 39.6; DB 3; Length 5433;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 242 TATATACATACATGATGATATATTTCTGTGTGTTTGTGCATATATAACCTCAACA 301
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5114 TATATATGATACAAACGATATATTTATGTCTATTTCAAAATAATATACATATTTCTAAA 5173
QY 302 CTATTATTAAATGCAATCCTATATTTCTTAGGTATAGAAGTTGATGATATA 351
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5174 GAATTATAAAATATTATGCAATTTTCAAAATAATAAAATTTTGTATTA 5223

RESULT 13
US-08-874-563-5/c
; Sequence 5, Application US/08874563
; Patent No. 6002071
; GENERAL INFORMATION:
; APPLICANT: CHAPPELL, JOSEPH
; APPLICANT: NEWMAN, JEFFREY D.
; APPLICANT: YIN, SHAOHUI
; TITLE OF INVENTION: TRANSCRIPTIONAL SILENCING
; TITLE OF INVENTION: ELEMENTS AND THEIR BINDING FACTORS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,563
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,087
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark, Esq.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 07678/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1368 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-874-563-5

Query Match
Best Local Similarity 7.9%; Score 39.4; DB 3; Length 1368;
Matches 44; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

[illegible][illegible]

Search completed: July 4, 2003, 19:02:49  
Job time : 36.9534 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 16:40:54 ; Search time 111.851 Seconds

(without alignments)

6978.003 Million cell updates/sec

Title: US-10-083-853b-2\_COPY\_1\_500

Perfect score: 500

Sequence: 1 gtaatgttaagaagcctca.....actaaagdttttcaggct 500

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PTCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	500	100.0	1934	9	US-10-086-623-5	Sequence 5, Appli
C 2	500	100.0	1934	9	US-10-260-539-5	Sequence 5, Appli
C 3	500	100.0	2253	9	US-10-086-623-7	Sequence 7, Appli
C 4	500	100.0	2253	9	US-10-260-539-7	Sequence 7, Appli
C 5	500	100.0	3718	9	US-10-258-557-3	Sequence 3, Appli
C 6	500	100.0	3736	9	US-10-258-557-1	Sequence 1, Appli
C 7	500	100.0	29921	9	US-10-083-853-1	Sequence 1, Appli
C 8	499.6	99.9	4001	10	US-09-915-582-35	Sequence 35, Appli
C 9	498.4	99.7	3798	10	US-09-915-582-17	Sequence 17, Appli
C 10	497.6	99.5	2726	10	US-09-915-582-37	Sequence 37, Appli
C 11	482.8	96.6	1882	9	US-10-139-583-36	Sequence 36, Appli
C 12	482.8	96.6	1882	9	US-10-039-847A-1	Sequence 1, Appli
C 13	482.8	96.6	1882	9	US-10-226-559-1	Sequence 1, Appli
C 14	482.8	96.6	1882	9	US-10-274-638-1	Sequence 1, Appli
C 15	482.8	96.6	1882	10	US-09-808-972-1	Sequence 1, Appli
C 16	460	92.0	485	9	US-09-954-531-902	Sequence 902, App
C 17	131.8	26.4	360	9	US-10-086-623-1	Sequence 1, Appli
C 18	131.8	26.4	360	9	US-10-260-539-1	Sequence 1, Appli
C 19	74.2	14.8	1472	9	US-10-039-847A-3	Sequence 3, Appli

Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 13, Appli  
Sequence 13, Appli  
Sequence 13, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 805, App  
Sequence 1238, Ap  
Sequence 177, App  
Sequence 4, Appli  
Sequence 3355, Ap  
Sequence 6381, Ap  
Sequence 4, Appli  
Sequence 4637, Ap  
Sequence 5333, Ap  
Sequence 8, Appli  
Sequence 14, Appli  
Sequence 60, Appli  
Sequence 9, Appli  
Sequence 27, Appli

#### ALIGNMENTS

##### RESULT 1

US-10-086-623-5/C

Sequence 5, Application US/100866623

Patent No. US20020164710A1

GENERAL INFORMATION:

APPLICANT: ERIKSSON, Ulf

APPLICANT: AASE, Karin

APPLICANT: LI, Xuri

APPLICANT: PONTEN, Anica

APPLICANT: UUTELA, Marko

APPLICANT: ALITALO, Kari

APPLICANT: OESTMAN, Arne

APPLICANT: HELDIN, Carl-Henrik

TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES

FILE REFERENCE: 1064/44833C2

CURRENT APPLICATION NUMBER: US/10/086,623

PRIOR FILING DATE: 2000-03-04

PRIOR APPLICATION NUMBER: US 60/107,852

PRIOR FILING DATE: 1998-11-10

PRIOR APPLICATION NUMBER: US 60/113,997

PRIOR FILING DATE: 1998-12-28

PRIOR APPLICATION NUMBER: US 60/150,604

PRIOR FILING DATE: 1999-08-26

PRIOR APPLICATION NUMBER: US 60/157,108

PRIOR FILING DATE: 1999-10-04

PRIOR APPLICATION NUMBER: US 60/157,756

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: US 09/438,046

PRIOR FILING DATE: 1999-11-10

PRIOR APPLICATION NUMBER: US 09/691,200

PRIOR FILING DATE: 2000-10-19

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 1934

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(966)

OTHER INFORMATION:

US-10-086-623-5

Query Match 100.0%; Score 500; DB 9; Length 1934;  
Best Local Similarity 100.0%; Pred. No. 3.8e-99;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTTCTTTTAAGAGA 60  
DB 1493 GTATATGTAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTTCTTTTAAGAGA 1434

QY 61 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATATATAATTTATCTCTCTAAATTTT 120  
DB 1433 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATATATAATTTATCTCTCTAAATTTT 1374

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGATTTTAAAGAGTCT 180  
DB 1373 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGATTTTAAAGAGTCT 1314

QY 181 AACTCAACATATGTAAAGCTCTGGTGACCTGGTATATATACCAAAAAAATTTTATGAT 240  
DB 1313 AACTCAACATATGTAAAGCTCTGGTGACCTGGTATATATACCAAAAAAATTTTATGAT 1254

QY 241 CTATATACATAGACATGAATATATTTCTGTGTCTTGTGATATATACCAAAATTTTATGAT 300  
DB 1253 CTATATACATAGACATGAATATATTTCTGTGTCTTGTGATATATACCAAAATTTTATGAT 1194

QY 301 ACTATTATTAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 360  
DB 1193 ACTATTATTAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 1134

QY 361 TTGCCATGGCATTAACAAGCAAGGCTGACACTCAGCAACACCTTGTGTTTCATTCATG 420  
DB 1133 TTGCCATGGCATTAACAAGCAAGGCTGACACTCAGCAACACCTTGTGTTTCATTCATG 1074

QY 421 CAGCTAGTAGTAAGTTTGGTGTCTGTAGGAAAGGCTCTTATCTCACCTCCCTTAA 480  
DB 1073 CAGCTAGTAGTAAGTTTGGTGTCTGTAGGAAAGGCTCTTATCTCACCTCCCTTAA 1014

QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
DB 1013 ACTAAAGGTTCTTTTCAGGCT 994

RESULT 2  
US-10-260-539-5/c  
Sequence 5, Application US/10260539  
Publication No. US20030073637A1  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: AASE, Karin  
APPLICANT: LI, Xuri  
APPLICANT: PONTEN, Annica  
APPLICANT: UUTELA, Marko  
APPLICANT: ALITALO, Kari  
APPLICANT: OESTMAN, Arne  
TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES TH  
FILE REFERENCE: 1064/44833C2  
CURRENT FILING DATE: 2002-10-01  
PRIORITY FILING DATE: 2002-10-01  
PRIORITY FILING DATE: 2000-03-04  
PRIORITY FILING DATE: 1998-11-10  
PRIORITY FILING DATE: 1998-12-28  
PRIORITY FILING DATE: 1999-08-26  
PRIORITY FILING DATE: 1999-08-26  
PRIORITY FILING DATE: 1999-10-04  
PRIORITY FILING DATE: 1999-10-05  
PRIORITY FILING DATE: 1999-11-10

PRIOR APPLICATION NUMBER: US 09/691,200  
PRIOR FILING DATE: 2000-10-19  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5  
LENGTH: 1934  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(966)  
OTHER INFORMATION:  
US-10-260-539-5

Query Match 100.0%; Score 500; DB 9; Length 1934;  
Best Local Similarity 100.0%; Pred. No. 3.8e-99;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTTCTTTTAAGAGA 60  
DB 1493 GTATATGTAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTTCTTTTAAGAGA 1434

QY 61 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATATATAATTTATCTCTCTAAATTTT 120  
DB 1433 GCAAGATTCAAAATTTGTTTGTCTTCAAAATTTAAATATATAATTTATCTCTCTAAATTTT 1374

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGATTTTAAAGAGTCT 180  
DB 1373 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGATTTTAAAGAGTCT 1314

QY 181 AACTCAACATATGTAAAGCTCTGGTGACCTGGTATATATACCAAAAAAATTTTATGAT 240  
DB 1313 AACTCAACATATGTAAAGCTCTGGTGACCTGGTATATATACCAAAAAAATTTTATGAT 1254

QY 241 CTATATACATAGACATGAATATATTTCTGTGTCTTGTGATATATACCAAAATTTTATGAT 300  
DB 1253 CTATATACATAGACATGAATATATTTCTGTGTCTTGTGATATATACCAAAATTTTATGAT 1194

QY 301 ACTATTATTAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 360  
DB 1193 ACTATTATTAATGCAATCCTATATTTCTAGGTATAGAAGTTGATGATATACCTTTCTAC 1134

QY 361 TTGCCATGGCATTAACAAGCAAGGCTGACACTCAGCAACACCTTGTGTTTCATTCATG 420  
DB 1133 TTGCCATGGCATTAACAAGCAAGGCTGACACTCAGCAACACCTTGTGTTTCATTCATG 1074

QY 421 CAGCTAGTAGTAAGTTTGGTGTCTGTAGGAAAGGCTCTTATCTCACCTCCCTTAA 480  
DB 1073 CAGCTAGTAGTAAGTTTGGTGTCTGTAGGAAAGGCTCTTATCTCACCTCCCTTAA 1014

QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
DB 1013 ACTAAAGGTTCTTTTCAGGCT 994

RESULT 3  
US-10-086-623-7/c  
Sequence 7, Application US/10086623  
Patent No. US20020164710A1  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: AASE, Karin  
APPLICANT: LI, Xuri  
APPLICANT: PONTEN, Annica  
APPLICANT: UUTELA, Marko  
APPLICANT: ALITALO, Kari  
APPLICANT: OESTMAN, Arne  
TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES  
FILE REFERENCE: 1064/44833C2  
CURRENT FILING DATE: 2000-03-04  
PRIORITY FILING DATE: 2000-03-04  
PRIORITY FILING DATE: 2000-03-04

~~RESULT 4~~  
US-10-260-539-7/c  
; Sequence 7, Application US/10260539  
; Publication No. US20030073637A1  
; GENERAL INFORMATION:  
; APPLICANT: ERIKSSON, Ulf  
; APPLICANT: AASE, Karin

Qy	1	GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTTAAGAGA	60
Db	1812	GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAAGATGCTTCTTTTAAGAGA	1753
Qy	61	GCAAGATTCAAATTTGTTTGGTTTCAAATTTAAATAAAATTTATCTCCCTAAATTTT	120
Db	1752	GCAAGATTCAAATTTGTTTGGTTTCAAATTTAAATAAAATTTATCTCCCTAAATTTT	1693
Qy	121	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGTATTTTAAGAGTCT	180
Db	1692	CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGTATTTTAAGAGTCT	1633
Qy	181	AACCTCAACATATGTAGAGCTCTGGTGACCTGGTTATATATACCAAAAACATTTTGAT	240
Db	1632	AACCTCAACATATGTAGAGCTCTGGTGACCTGGTTATATATACCAAAAACATTTTGAT	1573
Qy	241	CTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGTCATATATAACCTCAAAAC	300
Db	1572	CTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGTCATATATAACCTCAAAAC	1513
Qy	301	ACTATTATTAATGCAATCCATATTTCTTAGGTATAGAAGTTGATGATATACCTTCTAC	360
Db	1512	ACTATTATTAATGCAATCCATATTTCTTAGGTATAGAAGTTGATGATATACCTTCTAC	1453
Qy	361	TTGCGATGGCATTAACAAAACCAAGGCTGAGACTCAGCAACACCTTGTTGTTTCATTGCAATTG	420
Db	1452	TTGCGATGGCATTAACAAAACCAAGGCTGAGACTCAGCAACACCTTGTTGTTTCATTGCAATTG	1393
Qy	421	CAGGCTAGTAGTAAGTTTGGTTGCTGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA	480
Db	1392	CAGGCTAGTAGTAAGTTTGGTTGCTGTAGGAAAAGGTCCTTATCTCACCTCCCTTAA	1333



1000  
900  
800  
700  
600  
500  
400  
300  
200  
100  
0

✓  
RESULT 7

US-10-083-853-1  
; Sequence 1, Application US/10083853  
; Patent No. US20020164709A1  
; GENERAL INFORMATION:  
; APPLICANT: Affymetrix, Inc  
; APPLICANT: Shigeta, Ron T  
; APPLICANT: Siani-Rose, Michael A  
; TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein  
; FILE REFERENCE: 3385.1  
; CURRENT APPLICATION NUMBER: US/10/083,853  
; PRIOR FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: USSN 60/272,663  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 29921  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-083-853-1

Query Match 100.0%; Score 500; DB 9; Length 29921;  
Best Local Similarity 100.0%; Pred. No. 1.2e-98;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
|||||  
DB 1 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
|||||

QY 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 120  
|||||  
DB 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 120  
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGCAAGAGATTTTAAGAGTCT 180  
|||||  
DB 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGCAAGAGATTTTAAGAGTCT 180  
|||||

QY 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTATATATACCAAAAAACATTTGAT 240  
|||||  
DB 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTATATATACCAAAAAACATTTGAT 240  
|||||

QY 241 CTATATACATAGACATGAATATTTCTGTGTGTTCTGTGCATATATAACCTCAAC 300  
|||||  
DB 241 CTATATACATAGACATGAATATTTCTGTGTGTTCTGTGCATATATAACCTCAAC 300  
|||||

QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360  
|||||  
DB 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360  
|||||

QY 361 TTGCATGGCATTTACAAGCAAGCTCAGACTCAGCAACACCTTGTTCTATTGCAATTG 420  
|||||  
DB 361 TTGCATGGCATTTACAAGCAAGCTCAGACTCAGCAACACCTTGTTCTATTGCAATTG 420  
|||||

QY 421 CAGGCTAGTAGTAAGTTGGTTGGTGGTAGGAAAGGCTCTCTATCTCACCCCTCTTAA 480  
|||||  
DB 421 CAGGCTAGTAGTAAGTTGGTTGGTGGTAGGAAAGGCTCTCTATCTCACCCCTCTTAA 480  
|||||

QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
|||||  
DB 481 ACTAAAGGTTCTTTTCAGGCT 500  
|||||

## RESULT 8

US-09-915-582-35/c  
; Sequence 35, Application US/09915582  
; Patent No. US20020120103A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 17 Human Secreted Proteins  
; FILE REFERENCE: PS723P1

; CURRENT APPLICATION NUMBER: US/09/915,582  
; CURRENT FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: PCT/US01/01431  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/231,968  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 97  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 4001  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-915-582-35

Query Match 99.9%; Score 499.6; DB 10; Length 4001;  
Best Local Similarity 99.8%; Pred. No. 6.2e-99;  
Matches 499; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
|||||  
DB 1999 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 1940  
|||||

QY 61 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 120  
|||||  
DB 1939 GCAAGATTCAAAATGTTTGTGTTTCAAAATTTAAATTTATCTCTCTAAATTTT 1880  
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGCCAAAGATTTTAAGAGTCT 180  
|||||  
DB 1879 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGCCAAAGATTTTAAGAGTCT 1820  
|||||

QY 181 AACTCAACATATGTAAGCTCTGTGTACCTGGTATATATACCAAAAAACATTTGAT 240  
|||||  
DB 1819 AACTCAACATATGTAAGCTCTGTGTACCTGGTATATATACCAAAAAACATTTGAT 1760  
|||||

QY 241 CTATATACATAGACATGAATATTTCTGTGTGTTCTGTGCATATATAACCTCAAC 300  
|||||  
DB 1759 CTATATACATAGACATGAATATTTCTGTGTGTTCTGTGCATATATAACCTCAAC 1700  
|||||

QY 301 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 360  
|||||  
DB 1699 ACTATTATTAATGCAATCCCTATATTTCTAGGTATAGAGTTGATATACCTTTCTAC 1640  
|||||

QY 361 TTGCATGGCATTTACAAGCAAGCTCAGACTCAGCAACACCTTGTTCTATTGCAATTG 420  
|||||  
DB 1639 TTGCATGGCATTTACAAGCAAGCTCAGACTCAGCAACACCTTGTTCTATTGCAATTG 1580  
|||||

QY 421 CAGGCTAGTAGTAAGTTGGTTGGTGGTAGGAAAGGCTCTCTATCTCACCCCTCTTAA 480  
|||||  
DB 1579 CAGGCTAGTAGTAAGTTGGTTGGTGGTAGGAAAGGCTCTCTATCTCACCCCTCTTAA 1520  
|||||

QY 481 ACTAAAGGTTCTTTTCAGGCT 500  
|||||  
DB 1519 ACTAAAGGTTCTTTTCAGGCT 1500  
|||||

RESULT 9  
US-09-915-582-17/c  
; Sequence 17, Application US/09915582  
; Patent No. US20020120103A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 17 Human Secreted Proteins  
; FILE REFERENCE: PS723P1  
; CURRENT APPLICATION NUMBER: US/09/915,582  
; CURRENT FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: PCT/US01/01431  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31



```
; LOCATION: (226)...(1338)
US-10-139-583-36
Query Match          96.6%; Score 482.8; DB 9; Length 1882;
Best Local Similarity 98.8%; Pred. No. 2e-95;
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
DB 1866 GTATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 1807

QY 61 GCAAGATTCAAATTTGTTTGCATTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 120
DB 1806 GCAAGATTCAAATTTGTTTGCATTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 1747

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
DB 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 1687

QY 181 AACTCAAAACATATGTAAGCTCTGGTGTACCTGGTTATATATATACC-----AAAAAAACATT 236
DB 1686 AACTCAAAACATATGTAAGCTCTGGTGTACCTGGTTATATATATACC-----AAAAAAACATT 1627

QY 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATACCTTC 296
DB 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATACCTTC 1567

QY 297 AAACACTATTATTAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTT 356
DB 1566 AAACACTATTATTAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTT 1507

QY 357 CTACTTGGCATGGCATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGC 416
DB 1506 CTACTTGGCATGGCATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGC 1447

QY 417 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 476
DB 1446 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 1387

QY 477 TTAATACTAAAGGTTCTTTTCAGGCT 500
DB 1386 TTAATACTAAAGGTTCTTTTCAGGCT 1363

RESULT 12
US-10-039-847A-1/c
; Sequence 1, Application US/10039847A
; Publication No. US20020183273A1
; GENERAL INFORMATION:
; APPLICANT: Hart, Charles E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IMPROVING
; TITLE OF INVENTION: KIDNEY FUNCTION
; FILE REFERENCE: 00-100
; CURRENT APPLICATION NUMBER: US/10/039,847A
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/244,479
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)...(1338)
US-10-039-847A-1

Query Match          96.6%; Score 482.8; DB 9; Length 1882;
Best Local Similarity 98.8%; Pred. No. 2e-95;
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
DB 1866 GTATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 1807

QY 61 GCAAGATTCAAATTTGTTTGCATTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 120
DB 1806 GCAAGATTCAAATTTGTTTGCATTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 1747

QY 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 180
DB 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGAGATTTTAAAGAGTCT 1687

QY 181 AACTCAAAACATATGTAAGCTCTGGTGTACCTGGTTATATATATACC-----AAAAAAACATT 236
DB 1686 AACTCAAAACATATGTAAGCTCTGGTGTACCTGGTTATATATATACC-----AAAAAAACATT 1627

QY 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATACCTTC 296
DB 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATACCTTC 1567

QY 297 AAACACTATTATTAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTT 356
DB 1566 AAACACTATTATTAATGCAATCCCTATATTTCTTAGGTATAGAGTTGATGATATACCTTT 1507

QY 357 CTACTTGGCATGGCATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGC 416
DB 1506 CTACTTGGCATGGCATTAAACAAAGCAAGCTGAGACTCAGCAACCACTTGTGTTCATTGC 1447

QY 417 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 476
DB 1446 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 1387

QY 477 TTAATACTAAAGGTTCTTTTCAGGCT 500
DB 1386 TTAATACTAAAGGTTCTTTTCAGGCT 1363

RESULT 13
US-10-226-559-1/c
; Sequence 1, Application US/10226559
; Publication No. US20030105015A1
; GENERAL INFORMATION:
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,
; TITLE OF INVENTION: LIGAMENT AND CARTILAGE USING ZVBSF4
; FILE REFERENCE: 00-28
; CURRENT APPLICATION NUMBER: US/10/226,559
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: US/09/540,224
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 60/180,169
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)...(1338)
US-10-226-559-1

Query Match          96.6%; Score 482.8; DB 9; Length 1882;
Best Local Similarity 98.8%; Pred. No. 2e-95;
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 GATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 60
DB 1866 GTATATGTAAGAAAGCCCTCATCTTTTGATTTTAAATATACAAGATGCTTTCTTTAAAGAGA 1807

QY 61 GCAAGATTCAAATTTGTTTGCATTTTCAAAATTTAAATTAATTTATCTCTAAATTTT 120
```

Db 1806 GCAAGATTCACAAATGTTTCTGCTGTTTCAAAATTTAAATAATTTATCTCTAAATTTT 1747  
Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 180  
Db 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 1887  
Qy 181 AACTCAACATATATGTAAGCTCTGGTGACCTGGTGATATATACCCCTTATATATACCC 236  
Db 1686 AACTCAACATATGTAAGCTCTGGTGACCTGGTGATATATATACCCCTTATATATACCC 1627  
Qy 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 296  
Db 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 1567  
Qy 297 AACACATATTTAAATGCAATCCTATATTTCTTAGGTATAGAGTTTGTATATACCTTT 356  
Db 1566 AACACATATTTAAATGCAATCCTATATTTCTTAGGTATAGAGTTTGTATATACCTTT 1507  
Qy 357 CTACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGC 416  
Db 1506 CTACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGC 1447  
Qy 417 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 476  
Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 1387  
Qy 477 TTAACCTAAAGGTTCTTTTCAGGCT 500  
Db 1386 TTAACCTAAAGGTTCTTTTCAGGCT 1363

## RESULT 14

US-10-274-638-1/c  
; Sequence 1, Application US/10274638  
; Publication No. US20030109000A1  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Margaret D.  
; APPLICANT: Fox, Brian A.  
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS  
; TITLE OF INVENTION: AND METHODS FOR PRODUCING IT  
; FILE REFERENCE: 01-30  
; CURRENT APPLICATION NUMBER: US/10/274,638  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 60/346,117  
; PRIOR FILING DATE: 2001-10-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1882  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)...(1338)  
US-10-274-638-1

Query Match 96.6%; Score 482.8; DB 9; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 2e-95;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
Qy 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTTCTTTAAGAGA 60  
Db 1866 GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTTCTTTAAGAGA 1807  
Qy 61 GCAAGATTCACAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCTCAAAATTTT 120  
Db 1806 GCAAGATTCACAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCTCAAAATTTT 1747  
Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 180  
Db 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 1687

Qy 181 AACTCAACATATGTAAGCTCTGGTGACCTGGTGATATATATACCCCTTATATATACCC 236  
Db 1686 AACTCAACATATGTAAGCTCTGGTGACCTGGTGATATATATATACCCCTTATATATACCC 1627  
Qy 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 296  
Db 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTTTGTGCATATATAACCTC 1567  
Qy 297 AACACATATTTAAATGCAATCCTATATTTCTTAGGTATAGAGTTTGTATATACCTTT 356  
Db 1566 AACACATATTTAAATGCAATCCTATATTTCTTAGGTATAGAGTTTGTATATACCTTT 1507  
Qy 357 CTACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGC 416  
Db 1506 CTACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGC 1447  
Qy 417 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 476  
Db 1446 ATTGCAGGCTAGTAGTAAGTTTGGTGTGCTGGTAGGAAAGGCTCTTATCTCACCCCTCC 1387  
Qy 477 TTAACCTAAAGGTTCTTTTCAGGCT 500  
Db 1386 TTAACCTAAAGGTTCTTTTCAGGCT 1363

## RESULT 15

US-09-808-972-1/c  
; Sequence 1, Application US/09808972  
; Patent No. US20020064832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hart, Charles E.  
; APPLICANT: Topouzis, Stavros  
; APPLICANT: Gilbertson, Debra G.  
; TITLE OF INVENTION: METHOD OF TREATING FIBROPROLIFERATIVE  
; TITLE OF INVENTION: DISORDERS  
; FILE REFERENCE: 00-79  
; CURRENT APPLICATION NUMBER: US/09/808,972  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: US 60/235,295  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US 09/564,595  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: US 60/180,169  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/164,463  
; PRIOR FILING DATE: 1999-11-10  
; PRIOR APPLICATION NUMBER: US 60/132,250  
; PRIOR FILING DATE: 1999-05-03  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1882  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)...(1338)  
US-09-808-972-1

Query Match 96.6%; Score 482.8; DB 10; Length 1882;  
Best Local Similarity 98.8%; Pred. No. 2e-95;  
Matches 498; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
Qy 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTTCTTTAAGAGA 60  
Db 1866 GTATATGTAAGAAGCCTCATCTTTTGATTTTTTAATATACAAGATGCTTTCTTTAAGAGA 1807  
Qy 61 GCAAGATTCACAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCTCAAAATTTT 120  
Db 1806 GCAAGATTCACAAATGTTTGTGTTTCAAAATTTAAATAATTTATCTCTCAAAATTTT 1747  
Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGAGTCT 180

Db 1746 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 1687  
QY 181 AACTCAAAACATATGTAAGCTCTGTGTACCTGGTTATATATACC----AAAAAAACATT 236  
|||||  
Db 1686 AACTCAAAACATATGTAACCTCTGTGTACCTGGTTATATATACCAAAAAAAACATT 1627  
|||||  
QY 237 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTC 296  
|||||  
Db 1626 TGATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTC 1567  
|||||  
QY 297 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 356  
|||||  
Db 1566 AAACACTATTTAAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTT 1507  
|||||  
QY 357 CTACTTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACACCTTGTGTTCAATTGC 416  
|||||  
Db 1506 CTACTTGCCATGGCATTAAACAAGCAAGGCTGAGACTCAGCAACACCTTGTGTTCAATTGC 1447  
|||||  
QY 417 ATTGCAGGCTAGTAGTAGTTTGGTTGCTGTGTAGGAAAGGTTCTCTTATCTCACCCCTCC 476  
|||||  
Db 1446 ATTGCAGGCTAGTAGTAGTTTGGTTGCTGTGTAGGAAAGGTTCTCTTATCTCACCCCTCC 1387  
|||||  
QY 477 TTAAGCTAAAGGTTCTTTCAGGCT 500  
|||||  
Db 1386 TTAAGCTAAAGGTTCTTTCAGGCT 1363  
|||||

Search completed: July 4, 2003, 19:36:11  
Job time : 113.851 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:08:57 ; Search time 1508.67 Seconds  
(without alignments)  
9664.474 Million cell updates/sec

File: us-10-083-853b-2\_copy\_29421\_29921  
Perfect score: 501  
Sequence: 1 ctttggtgtttgacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database 8 GenEmbl80

- 1: gb-ba.\*
- 2: gb-htg.\*
- 3: gb-in.\*
- 4: gb-om.\*
- 5: gb-lov.\*
- 6: gb-pat.\*
- 7: gb-ph.\*
- 8: gb-pl.\*
- 9: gb-pr.\*
- 10: gb-ro.\*
- 11: gb-sts.\*
- 12: gb-sy.\*
- 13: gb-un.\*
- 14: gb-vi.\*
- 15: em-ba.\*
- 16: em-fun.\*
- 17: em-hum.\*
- 18: em-in.\*
- 19: em-mu.\*
- 20: em-or.\*
- 21: em-ov.\*
- 22: em-pat.\*
- 23: em-ph.\*
- 24: em-pl.\*
- 25: em-ro.\*
- 26: em-sts.\*
- 27: em-un.\*
- 28: em-vi.\*
- 29: em-htg-hum.\*
- 30: em-htg-inv.\*
- 31: em-htg-other.\*
- 32: em-htg-mus.\*
- 33: em-htg-pln.\*
- 34: em-htg-rod.\*
- 35: em-htg-mam.\*
- 36: em-htg-vrt.\*
- 37: em-sy.\*
- 38: em-hgo-hum.\*
- 39: em-hgo-mus.\*
- 40: em-hgo-other.\*
- 41: em-hgo-other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	180648	9	AP003043	AP003043 Homo sapi
2	501	100.0	182403	2	AC024052	AC024052 Homo sapi
3	480.2	95.8	46442	9	AL591394	AL591394 Human DNA
4	478.6	95.5	47364	9	AC008809	AC008809 Homo sapi
5	478.6	95.5	206163	9	AC099519	AC099519 Homo sapi
6	478.6	95.5	206865	2	AC027027	AC027027 Homo sapi
7	477	95.2	94727	9	AL591034	AL591034 Human DNA
8	477	95.2	149041	9	AC044782	AC044782 Homo sapi
9	477	95.2	154076	2	AC023273	AC023273 Homo sapi
10	477	95.2	162841	9	AC099504	AC099504 Homo sapi
11	477	95.2	169161	2	AC019053	AC019053 Homo sapi
12	477	95.2	171490	2	AC068209	AC068209 Homo sapi
13	477	95.2	176630	2	AC010684	AC010684 Homo sapi
14	477	95.2	179789	9	AC069843	AC069843 Homo sapi
15	477	95.2	184000	9	AC087774	AC087774 Homo sapi
16	475.4	94.9	3994	9	AC111191	AC111191 Homo sapi
17	475.4	94.9	41260	9	AC114309	AC114309 Homo sapi
18	475.4	94.9	116800	9	AL391379	AL391379 Human DNA
19	475.4	94.9	142391	2	AC092507	AC092507 Homo sapi
20	475.4	94.9	147423	9	AC098487	AC098487 Homo sapi
21	475.4	94.9	150572	9	CNS01RIC	AL163151 Human chr
22	475.4	94.9	151018	9	AC096756	AC096756 Homo sapi
23	475.4	94.9	152741	2	AC021606	AC021606 Homo sapi
24	475.4	94.9	156485	9	AC007422	AC007422 Homo sapi
25	475.4	94.9	163673	2	AC027598	AC027598 Homo sapi
26	475.4	94.9	165295	2	AC068486	AC068486 Homo sapi
27	475.4	94.9	165714	9	AC016919	AC016919 Homo sapi
28	475.4	94.9	168063	9	AC080051	AC080051 Homo sapi
29	475.4	94.9	172543	9	AC106714	AC106714 Homo sapi
30	475.4	94.9	172805	9	AC023134	AC023134 Homo sapi
31	475.4	94.9	173288	9	AC104167	AC104167 Homo sapi
32	475.4	94.9	176624	2	AC104394	AC104394 Homo sapi
33	475.4	94.9	176903	9	AC008501	AC008501 Homo sapi
34	475.4	94.9	178402	2	AC090508	AC090508 Homo sapi
35	475.4	94.9	178780	9	AF216667	AF216667 Homo sapi
36	475.4	94.9	180339	9	AC099049	AC099049 Homo sapi
37	475.4	94.9	181236	2	AC022757	AC022757 Homo sapi
38	475.4	94.9	183437	9	AC068933	AC068933 Homo sapi
39	475.4	94.9	189053	2	AC016015	AC016015 Homo sapi
40	475.4	94.9	190195	9	AC016954	AC016954 Homo sapi
41	475.4	94.9	192586	9	AC018437	AC018437 Homo sapi
42	475.4	94.9	194867	2	AC023152	AC023152 Homo sapi
43	475.4	94.9	195931	2	AC128693	AC128693 Homo sapi
44	475.4	94.9	196660	9	AC007000	AC007000 Homo sapi
45	475.4	94.9	200466	2	AL845433	AL845433 Homo sapi

ALIGNMENTS

RESULT 1  
AP003043  
LOCUS  
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-617B3, complete sequence.  
ACCESSION AP003043  
VERSION AP003043.2 GI:15637094  
KEYWORDS HTG.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE  
1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
linear PR1 18-SEP-2001

TITLE Homo sapiens genomic DNA  
JOURNAL Published Only in Database (2000)  
REFERENCE 2 (bases 1 to 180648)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT On Sep 17, 2001 this sequence version replaced gi:11862945.  
FEATURES  
source  
1..180648  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q"  
/clone="RP11-617B3"  
BASE COUNT 56234 a 35375 c 33753 g 55286 t  
ORIGIN

Query Match 100.0%; Score 501; DB 9; Length 180648;  
Best Local Similarity 100.0%; Pred. No. 5e-112;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTTTGGTGTAGACATCAAGTCCCTGCGCCATGCTATGCTGAATGGTATGGCTG 60  
Db CTTTGGTGTAGACATCAAGTCCCTGCGCCATGCTATGCTGAATGGTATGGCTG 69350  
Qy 61 GGTTCCTCTAGGATTTTATGGTTAGTCTAACATTTAAAGTCTTTAATCCATCTTG 120  
Db GGTTCCTCTAGGATTTTATGGTTAGTCTAACATTTAAAGTCTTTAATCCATCTTG 69410  
Qy 121 AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTCACGTTTCTACATAGGCTA 180  
Db AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTTCACGTTTCTACATAGGCTA 69470  
Qy 181 GCCAGTTTCTCAGCACCATTTAATAAGGAATCTTTCCCATGCTCTGTTTCT 240  
Db GCCAGTTTCTCAGCACCATTTAATAAGGAATCTTTCCCATGCTCTGTTTCT 69530  
Qy 241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATATTCTGAGGGCTCTGTT 300  
Db CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATATTCTGAGGGCTCTGTT 69590  
Qy 301 TGTTCATGTTGATATCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 360  
Db TGTTCATGTTGATATCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 69650  
Qy 361 CCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTGGCTTA 420  
Db CCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTGGCTTA 69710  
Qy 421 GGATGACTTGGCGATGTTGGCTCTTTTGGTTCCACATGAATCTTAAAGTAGTTTTT 480  
Db GGATGACTTGGCGATGTTGGCTCTTTTGGTTCCACATGAATCTTAAAGTAGTTTTT 69770  
Qy 481 CCAATCTGTGAAGAAGTAA 501  
Db CCAATCTGTGAAGAAGTAA 69791

RESULT 2  
AC024052 182403 bp DNA linear HTG 17-AUG-2000  
LOCUS Homo sapiens chromosome 11 clone RP11-617B3, WORKING DRAFT  
DEFINITION SEQUENCE, 7 unordered pieces.  
AC024052  
AC024052.3 GI:9838295  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 182403)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 182403)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Aug 17, 2000 this sequence version replaced gi:7109658.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site:http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0617B03  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: Plasmid; 0%  
Chemistry: Dye-primer ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 178574 bases at least Q40  
Consensus quality: 179480 bases at least Q30  
Consensus quality: 179940 bases at least Q20  
Insert size: 198000; agarose-fp  
Insert size: 183143; sum-of-contigs  
Quality coverage: 5.39 in Q20 bases; agarose-fp  
Quality coverage: 5.87 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 1666: contig of 1666 bp in length  
\* 1667 1766: gap of unknown length  
\* 1767 9133: contig of 7367 bp in length  
\* 9134 9233: gap of unknown length  
\* 9234 20663: contig of 11430 bp in length  
\* 20664 20763: gap of unknown length  
\* 20764 35927: contig of 15164 bp in length  
\* 35928 36028: gap of unknown length  
\* 36028 61610: contig of 25583 bp in length  
\* 61611 61710: gap of unknown length  
\* 61711 96167: contig of 34457 bp in length  
\* 96168 96268: gap of unknown length  
\* 96268 182403: contig of 86136 bp in length.  
FEATURES  
source  
1..182403  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-617B3"  
1..1666  
/note="assembly\_name:Contig2"  
1767..9133  
/note="assembly\_name:Contig3"  
clone\_end:SF6  
vector\_side:left  
9234..20663  
/note="assembly\_name:Contig4"  
20764..35927  
/note="assembly\_name:Contig5"  
36028..61610  
/note="assembly\_name:Contig6"

misc_feature		61711..96167		/note="assembly_name:Contig7"	
misc_feature		96268..182403		/note="assembly_name:Contig8	
		clone_end:77		vector_side:left"	
BASE COUNT		55923 a 35190 c 34519 g 56169 t		602 others	
ORIGIN					
Query Match		100.0%; Score 501; DB 2; Length 182403;			
Best Local Similarity		100.0%; Pred. No. 5e-112;			
Matches 501; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	CTTTTGGTGTTT	TAGACATGAAGTCCTTGGCCATGCCTATGCTCTGAATGGTATTCGCTG	60	
Db	165535	CTTTTGGTGTTT	TAGACATGAAGTCCTTGGCCATGCCTATGCTCTGAATGGTATTCGCTG	165594	
QY	61	GGTTTCTCTAGG	ATTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG	120	
Db	165595	GGTTTCTCTAGG	ATTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG	165654	
QY	121	AATTAATTTTGT	ATAAGTGTAAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA	180	
Db	165655	AATTAATTTTGT	ATAAGTGTAAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA	165714	
QY	181	GCCAGTTTCTC	AGCACCATTTATTAATAGGGAATCCTTTCCCATTCCTTGTGTTTCT	240	
Db	165715	GCCAGTTTCTC	AGCACCATTTATTAATAGGGAATCCTTTCCCATTCCTTGTGTTTCT	165774	
QY	241	CAGGTTTCTCA	AGATCAGATAGTTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTC	300	
Db	165775	CAGGTTTCTCA	AGATCAGATAGTTGTAGATATCGGCATTAATTTCTGAGGGCTCTGTC	165834	
QY	301	TGTTCCATTTG	GTTGATATCTCTGTTTGTGTACAGTACCAGTCTGTTTGTACTAGTAG	360	
Db	165835	TGTTCCATTTG	GTTGATATCTCTGTTTGTGTACAGTACCAGTCTGTTTGTACTAGTAG	165894	
QY	361	CTTGTAGTGT	AGTTTGAAGTCAGTATGATGCTTCCAGCTTTGTTTCTTTTGGCTTA	420	
Db	165895	CTTGTAGTGT	AGTTTGAAGTCAGTATGATGCTTCCAGCTTTGTTTCTTTTGGCTTA	165954	
QY	421	GGATTGACTT	GGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT	480	
Db	165955	GGATTGACTT	GGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT	166014	
QY	481	CCAATTCGTG	TGAAGAAAGTAA 501		
Db	166015	CCAATTCGTG	TGAAGAAAGTAA. 166035		
RESULT 3		AL591394/c			
LOCUS		46442 bp DNA linear PRI 27-OCT-2001			
DEFINITION		Human DNA sequence from clone Rp11-1145822 on chromosome			
		Xp11.21-11.23, complete sequence.			
ACCESSION		AL591394			
VERSION		AL591394.11			
KEYWORDS		HTG.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL		Whitehead S.			
TITLE		Direct Submission			
COMMENT		Submitted (27-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk			
		requests: clonerequests@sanger.ac.uk			
		On Oct 30, 2001 this sequence version replaced gi:15723107.			
		During difference assembly data is compared from overlapping clones.			
		Where differences are found these are annotated as variations			
		together with a note of the overlapping clone name. Note that the			
		variation annotation may not be found in the sequence submission			

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.			
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at			
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at			
http://www.sanger.ac.uk/HGP/ChrX			
Rp11-1145822 is from the library RPCI-11.4 constructed by the group of P. de Jong. For further details see			
http://www.chori.org/bacpac/home.htm			
VECTOR: pBAC3.6			
IMPORTANT: This sequence is not the entire insert of clone Rp11-1145822. It may be shorter because we sequenced overlapping sections only once, except for a short overlap.			
The true left end of clone Rp1-54820 is at 44443 in this sequence.			
The true right end of clone Rp3-393P12 is at 2000 in this sequence.			
FEATURES			
Location/Qualifiers	Source		
1..46442	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="X"		
	/map="p11.21-11.23"		
	/clone="Rp11-1145822"		
	/clone_lib="RPCI-11.4"		
BASE COUNT	13526 a 9133 c 9396 g 14387 t		
ORIGIN			
Query Match	95.8%; Score 480.2; DB 9; Length 46442;		
Best Local Similarity	97.4%; Pred. No. 6.6e-107;		
Matches 488; Conservative	0; Mismatches 13; Indels 0; Gaps 0;		
QY	1 CTTTGGTGTTT	TAGACATGAAGTCCTTGGCCCATGCCTATGCTCTGAATGGTATTCGCTG	60
Db	41112 CTTTGGTGTTT	TAGACATGAAGTCCTTGGCCCGTCTATGCTCCTGAATGGTATTCGCTA	4105
QY	61 GGTTCCTCT	CAGGATTTTATGGTTTATAGCTTAACATTTAAGCTTTAATCCATCTTG	120
Db	41052 AGTTTCTCT	CAGGATTTTATGGTTTATAGCTTAACATTTAAGCTTTAATCCATCTTG	4099
QY	121 AATTAATTTT	TGTAAGTGTAAGGAGGATCCAGTTTCAGCTTTCTACATAGGGCTA	180
Db	40992 AATTAATTTT	TGTAAGTGTAAGGAGGATCCAGTTTCAGCTTTCAACATATGGCTA	4093
QY	181 GCCAGTTTCT	CAGCACCATTTATTAATAGGGAATCCTTTCCCATTCCTTGTGTTTCT	240
Db	40932 GCCAGTTTCT	CAGCACCATTTATTAATAGGGAATCCTTTCCCATTCCTTGTGTTTCT	4087
QY	241 CAGGTTTCT	CAAGATCAGATAGTTGTAGATATGGGCATTTATTTCTGAGGGCTCTGTC	300
Db	40872 CAGGTTTCT	CAAGATCAGATAGTTGTAGATATGGGCATTTATTTCTGAGGGCTCTGTC	4081
QY	301 TGTCCATTTG	GTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTACTCTAG	360
Db	40812 TGTCCATTTG	GTCTATATCTCTGTTTGGTACCAGTACCATGTTTGGTACTCTAG	4075
QY	361 CCTTGTAGTG	TATGTTGAAGTCAGGTAGCATGATCCCTCCAGCTTTGTTTGGCTTA	420
Db	40752 CCTTGTAGTG	TATGTTGAAGTCAGGTAGCATGATCCCTCCAGCTTTGTTTGGCTTA	4069
QY	421 GGATTGACT	TGGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT	480
Db	40692 GGATTGACT	TGGCGATGGGCTCTTTTGGTCCATATGAACCTTTAAAGTAGTTTTT	4063

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>

Rp11-1145822 is from the library RPCI-11.4 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-1145822 it may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone Rp1-54B20 is at 44443 in this sequence.

The true right end of clone Rp3-393P12 is at 2000 in this sequence.

FEATURES

Location/Qualifiers

1..46442

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="X"

/map="p11.21-11.23"

/clone="Rp11-1145B22"

/clone\_lib="RPCI-11.4"

BASE COUNT 13526 a 9133 c 9396 g 14387 t

ORIGIN

Query Match 95.8%; Score 480.2; DB 9; Length 46442;

Best Local Similarity 97.4%; Pred. No. 6.6e-107;

Matches 488; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTAGACATGAAGTCCTTGGCCATGCCTATGCTCTGAATGTTATTCGCTG 60

DB 41112 CTTTGGTGTGTAGACATGAAGTCCTTGGCCATGCCTATGCTCTGAATGTTATTCGCTA 41053

QY 61 GGTTTCTCTAGGAGTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 120

DB 41052 AGTTTCTCTAGGAGTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 40993

QY 121 AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180

DB 40992 AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTCAGCTTTCAACATAGGCTA 40933

QY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCTTTCCCATTCCTTGGTGTGTTTCT 240

DB 40932 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCTTTCCCATTCCTTGGTGTGTTTCT 40873

QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATGCGGCATTAATTTCTGAGGCTCTGTC 300

DB 40872 CAGGTTTCTCAAGATCAGATAGTTGTAGATGCGGCATTAATTTCTGAGGCTCTGTC 40813

QY 301 TGTTCCATTTGGTTGATATCTCTGTTTGTGTACAGTACCAGTCTGTTTGTGTTACTAGTAG 360

DB 40812 TGTTCCATTTGGTTGATATCTCTGTTTGTGTACAGTACCAGTCTGTTTGTGTTACTAGTAG 40753

QY 361 CTTGTAGTGTAGTTTGAAGTCAGTATGATGCTTCCAGCTTTGTTTCTTTTGGCTA 420

DB 40752 CTTGTAGTGTAGTTTGAAGTCAGTATGATGCTTCCAGCTTTGTTTCTTTTGGCTA 40693

QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGACACTTAAAGTAGTTTTTT 480

DB 40692 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGACACTTAAAGTAGTTTTTT 40633

```
QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 40632 CCAATTCGTGAAGAAAGTCA 40612

RESULT 4
AC008809 47364 bp DNA linear PRI 28-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2093D7, complete sequence.
ACCESSION AC008809
VERSION AC008809.6 GI:14573671
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 47364)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 47364)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 47364)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (28-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jun 28, 2001 this sequence version replaced gi:13676970.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.
STS Content:
WI-2295 G02858.
FEATURES
source
1..47364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2093D7"
BASE COUNT 15105 a 8487 c 8493 g 15279 t
ORIGIN
Query Match 95.5%; Score 478.6; DB 9; Length 47364;
Best Local Similarity 97.2%; Pred. No. 1.6e-106;
Matches 487; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CTTTGTGTTGTAGACATGAAGTCCCTGGCCATGCCATGCTGCTGAATGGTATGCTG 60
Db 19625 CTTTGTGTTGTAGACATGAAGTCCCTGGCCATGCCATGCTGCTGAATGGTATGCTG 19684

QY 61 GGTTCCTCTCAGGATTTATGTTTGTAGTCTTAACATTAAGTCTTTAATCCATCTTG 120
Db 19685 GGTTCCTCTCAGGATTTATGTTTGTAGTCTTAACATTAAGTCTTTAATCCATCTTG 19744

QY 121 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180
Db 19745 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTATCAGCTTTCTACATAGGGCTA 19804

QY 181 GCCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCAATGCTGTTTCT 240
Db 19805 GCCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCAATGCTGTTTCT 19864

QY 241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTATTTCTGAGGGCTCTGTTTC 300
Db 19865 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTATTTCTGAGGGCTCTGTTTC 19924
```

---

```
QY 301 TGTCCATCGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTACTAG 360
Db 19925 CGTTCCATCGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTACTAG 19984

QY 361 CTTGTAGTCTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCCTTTGGCCTTA 420
Db 19985 CTTGTAGTCTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTCCTTTGGCCTTA 20044

QY 421 GGATTGACTTGGCGCATGCTGGCTCTCTTTTGGTTCACATGAACCTTTAAAGTAGTATTTT 480
Db 20045 GGATTGACTTGGCGCATGCTGGCTCTCTTTTGGTTCACATGAACCTTTAAAGTAGTATTTT 20104

QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 20105 CCAATTCGTGAAGAAAGTCA 20125

RESULT 5
AC099519 206163 bp DNA linear PRI 20-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-695O10, complete sequence.
ACCESSION AC099519
VERSION AC099519.2 GI:18767425
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 206163)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 206163)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 206163)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (20-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Feb 20, 2002 this sequence version replaced gi:16930935.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1..206163
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-695O10"
BASE COUNT 65125 a 37397 c 38034 g 65607 t
ORIGIN
Query Match 95.5%; Score 478.6; DB 9; Length 206163;
Best Local Similarity 97.3%; Pred. No. 1.4e-106;
Matches 487; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CTTTGTGTTGTAGACATGAAGTCCCTGGCCATGCCATGCTGCTGAATGGTATGCTG 60
Db 85553 CTTTGTGTTGTAGACATGAAGTCCCTGGCCATGCCATGCTGCTGAATGGTATGCTG 86494

QY 61 GGTTCCTCTCAGGATTTATGTTTGTAGTCTTAACATTAAGTCTTTAATCCATCTTG 120
Db 86493 GGTTCCTCTCAGGATTTATGTTTGTAGTCTTAACATTAAGTCTTTAATCCATCTTG 86434

QY 121 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180
Db 86433 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTATCAGCTTTCTACATAGGGCTA 86374
```

181 GCACGTTTCTCAGCACCATTATTAAGGAGGAAATCCTTTCCCATTCGTTGTTTCT 240  
|||||  
86373 GCCAGTTTCTCCAGCACCATTATTAAGGAGGAAATCCTTTCCCATTCGTTGTTTCT 86314  
|||||  
241 CAGGTTTTCGAAGATCAGATAGTGTGATAGATGCGGCAATATTTCTGAGGCTCTGTTC 300  
|||||  
86313 CAGGTTTTCGAAGATCAGATAGTGTGATAGATGCGGCAATATTTCTGAGGCTCTGTTC 86254  
|||||  
301 TCTTCCATTCGTTGATATCTCTGTTTGTACCATCAGCATCTTTTGGTTACTAGTAG 360  
|||||  
86253 CTTTCCATTCGTTGATATCTCTGTTTGTACCATCAGCATCTTTTGGTTACTAGTAG 86194  
|||||  
361 CTTTGTAGTGTAGTGTGAGTGTGATGATGCTTCCAGCTTTGTTTGGCTTGA 420  
|||||  
86193 CTTTGTAGTGTAGTGTGAGTGTGATGATGCTTCCAGCTTTGTTTGGCTTGA 86134  
|||||  
421 GGATTGACCTGCGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTGTTC 480  
|||||  
86133 GGATTGACCTGCGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTGTTC 86074  
|||||  
481 CCAATTCCTGTAAGAAAGTAA 501  
|||||  
86073 CCAATTCCTGTAAGAAAGTAA 86053  
|||||

## RESULT 6

AC027027/c AC027027 206865 bp DNA linear HTG 25-MAY-2000  
LOCUS Homo sapiens chromosome 5 clone RP11-695010 map 5, WORKING DRAFT  
DEFINITION SEQUENCE, 19 unordered pieces.  
AC027027  
AC027027.3 GI:8081300  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Aldredge,J., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G., Collins,S.,  
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczky,J.,  
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

## TITLE

JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On May 25, 2000 this sequence version replaced gi:7677916.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

FEATURES  
source

1..206865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5"  
/clone="RP11-695010"  
/clone\_lib="RPCI-11 Human Male BAC"  
1..1014  
/note="assembly\_fragment"

misc\_feature

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L8849  
Center clone name: 695\_O\_10  
----- Summary Statistics  
Sequencing vector: M13, M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 196102 bases at least Q40  
Consensus quality: 201507 bases at least Q30  
Consensus quality: 203540 bases at least Q20  
Insert size: 205000; agarose-fp  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.6 in Q20 bases; sum-of-contigs.  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1014: contig of 1014 bp in length  
1015 1114: gap of 100 bp  
1115 3783: contig of 2669 bp in length  
3784 3883: gap of 100 bp  
3884 6474: contig of 2591 bp in length  
6475 6574: gap of 100 bp  
6575 11183: contig of 4609 bp in length  
11184 11283: gap of 100 bp  
11284 16799: contig of 5516 bp in length  
16800 16899: gap of 100 bp  
16900 21561: contig of 4662 bp in length  
21562 21661: gap of 100 bp  
21662 27796: contig of 6135 bp in length  
27797 27896: gap of 100 bp  
27897 33998: contig of 6102 bp in length  
33999 34098: gap of 100 bp  
34099 41507: contig of 7409 bp in length  
41508 41607: gap of 100 bp  
41608 47859: contig of 6252 bp in length  
47860 47959: gap of 100 bp  
47960 56205: contig of 8246 bp in length  
56206 56305: gap of 100 bp  
56306 68561: contig of 12256 bp in length  
68562 68661: gap of 100 bp  
68662 81542: contig of 12881 bp in length  
81543 81642: gap of 100 bp  
81643 96660: contig of 15018 bp in length  
96661 96760: gap of 100 bp  
96761 114606: contig of 17846 bp in length  
114607 114706: gap of 100 bp  
114707 136359: contig of 21653 bp in length  
13660 136459: gap of 100 bp  
13660 156848: contig of 20189 bp in length  
156849 156748: gap of 100 bp  
156749 178778: contig of 22030 bp in length  
178779 178878: gap of 100 bp  
178879 206865: contig of 27987 bp in length.

Location/Qualifiers:  
1..206865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5"  
/clone="RP11-695010"  
/clone\_lib="RPCI-11 Human Male BAC"  
1..1014  
/note="assembly\_fragment"

misc\_feature

```
misc_feature 1115..3783
/note="assembly_fragment"
misc_feature 3884..6474
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 6575..11183
/note="assembly_fragment"
misc_feature 11284..16799
/note="assembly_fragment"
misc_feature 16900..21561
/note="assembly_fragment"
misc_feature 21662..27796
/note="assembly_fragment"
misc_feature 27897..33998
/note="assembly_fragment"
misc_feature 34099..41507
/note="assembly_fragment"
misc_feature 41608..47859
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature 47960..56205
/note="assembly_fragment"
misc_feature 56306..68561
/note="assembly_fragment"
misc_feature 68662..81542
/note="assembly_fragment"
misc_feature 81643..96660
/note="assembly_fragment"
misc_feature 96761..114606
/note="assembly_fragment"
misc_feature 114707..136359
/note="assembly_fragment"
misc_feature 136460..156648
/note="assembly_fragment"
misc_feature 156749..178778
/note="assembly_fragment"
misc_feature 178879..206865
/note="assembly_fragment"
BASE COUNT 64815 a 37677 c 37421 g 65147 t 1805 others
ORIGIN
Query Match 95.5%; Score 478.6; DB 2; Length 206865;
Best Local Similarity 97.2%; Pred. No. 1.4e-106;
Matches 487; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 CTTTGTGTTTGTAGATGAGTCCCTTGCCCATGCTATGCTGCTGAATGGTATGCTG 60
Db 71931 CTTTGTGTTTGTAGATGAGTCCCTTGCCCATGCTATGCTGCTGAATGGTATGCTG 71872
QY 61 GGTTCCTTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 120
Db 71871 GGTTCCTTCTAGGAGTTTATGTTTATGCTCTAACATTTAAGTCTTTAATCCATCTTG 71812
QY 121 AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
Db 71811 AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 71752
QY 181 GCCAGTTTCTCAGCACCAATTATTAATAGGGAATCTTCCCATGCTGCTTTTCT 240
Db 71751 GCCAGTTTCTCAGCACCAATTATTAATAGGGAATCTTCCCATGCTGCTTTTCT 71692
QY 241 CAGTTTCTCAAGATCAGATAGTTGTAGATATGCGCATATTTCTGAGGGCTCTGTC 300
Db 71691 CAGTTTCTCAAGATCAGATAGTTGTAGATATGCGCATATTTCTGAGGGCTCTGTC 71632
QY 301 TGTTCATTTGTTGATATCTGTTTGGTACAGTACCAAGTGTGTTTGGTACTGTAG 360
Db 71631 CGTTCCATTTGCTATATCTGTTTGGTACCAAGTGTGTTTGGTACTGTAG 71572
QY 361 CCTTGTAGTACTGTTTCAAGTCAGTAGCATGCTCCAGCTTGTCTTTTGGCTTA 420
Db 71571 CCTTGTAGTACTGTTTCAAGTCAGTAGCATGCTCCAGCTTGTCTTTTGGCTTA 71512
QY 421 GGATTGACTTGGCATGTGGCTCTTTTGGTTCACATGAACCTTTAAAGTACTTTTT 480
Db 71511 GGATTGACTTGGCATGTGGCTCTTTTGGTTCACATGAACCTTTAAAGTACTTTTT 71452
QY 481 CCAATTCCTGTGAAGAAAGTAA 501
Db 71451 CCAATTCCTGTGAAGAAAGTCA 71431
RESULT 7
AL591034 Human DNA sequence from clone RP11-96J19 on chromosome 6, complete
LOCUS AL591034 94727 bp DNA linear PRI 13-JUN-2001
DEFINITION sequence.
ACCESSION AL591034
VERSION AL591034.5 GI:14456374
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 94727)
AUTHORS Williams,S.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jun 15, 2001 this sequence version replaced gi:14270021.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-96J19 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-96J19 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-27K12 is at 92728 in this sequence.
The true right end of clone RP3-483K16 is at 2000 in this sequence.
FEATURES
source
1..94727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="6"
/clone="RP11-96J19"
/clone_lib="RPCI-11.1"
182..322
/note="L2 repeat: matches 2555..2707 of consensus"
repeat_region
443..588
/note="FLAM_C repeat: matches 1..143 of consensus"
repeat_region
613..732
```

```
repeat_region /note="Charlie5 repeat: matches 4. .116 of consensus"
repeat_region /note="Charlie5 repeat: matches 2241. .2391 of consensus"
repeat_region /note="L1MB4 repeat: matches 5628. .5767 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="THEIC-internal repeat: matches 500. .1580 of consensus"
repeat_region /note="AluSq repeat: matches 1. .293 of consensus"
repeat_region /note="THEIC-internal repeat: matches 3. .500 of consensus"
repeat_region /note="Alu repeat: matches 1. .311 of consensus"
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"
repeat_region /note="L1MB4 repeat: matches 5764. .6182 of consensus"
repeat_region /note="THEI-INTERNAL repeat: matches 752. .877 of consensus"
misc_feature /note="match: GSS: Em:A2576617"
repeat_region /note="85 copies 2 mer tt 67% conserved"
repeat_region /note="3 copies 49 mer 78% conserved"
repeat_region /note="AluSq repeat: matches 1. .309 of consensus"
repeat_region /note="AluSq repeat: matches 1. .307 of consensus"
repeat_region /note="FLAM_A repeat: matches 1. .133 of consensus"
repeat_region /note="AluSq repeat: matches 1. .308 of consensus"
repeat_region /note="MIR repeat: matches 71. .209 of consensus"
repeat_region /note="L2 repeat: matches 2640. .2749 of consensus"
misc_feature /note="match: GSS: Em:AQ106154"
repeat_region /note="AluSg/x repeat: matches 128. .290 of consensus"
misc_feature complement(13683. .14100)
misc_feature /note="match: GSS: Em:AQ265669"
misc_feature /note="match: GSS: Em:AQ683800"
repeat_region /note="match: STS: Em:G31644"
misc_feature /note="AluSx repeat: matches 4. .305 of consensus"
misc_feature complement(15538. .15933)
repeat_region /note="match: STS: Em:HSP50B05"
repeat_region /note="MIR repeat: matches 1. .262 of consensus"
repeat_region /note="L1MB4 repeat: matches 3896. .3961 of consensus"
repeat_region /note="AluSq repeat: matches 1. .305 of consensus"
repeat_region /note="AluSq repeat: matches 19. .293 of consensus"
repeat_region /note="L2 repeat: matches 2688. .2746 of consensus"
repeat_region /note="L1P5 repeat: matches 13. .2993 of consensus"
repeat_region /note="Alu repeat: matches 1. .302 of consensus"
repeat_region /note="L1P5 repeat: matches 2993. .6141 of consensus"
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
```

```
repeat_region 26039. .26068
repeat_region /note="20 copies 2 mer ca 85% conserved"
repeat_region 26137. .26166
repeat_region /note="7SLRNA repeat: matches 245. .274 of consensus"
repeat_region 26164. .26230
repeat_region /note="AluJ/FLAM repeat: matches 3. .71 of consensus"
repeat_region 26957. .27250
repeat_region /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 27958. .28203
repeat_region /note="L1M4 repeat: matches 5347. .5588 of consensus"
repeat_region 28207. .29139
repeat_region /note="MER21A repeat: matches 10. .902 of consensus"
repeat_region 29140. .29179
repeat_region /note="20 copies 2 mer ct 97% conserved"
repeat_region 29194. .29943
repeat_region /note="L1M10 repeat: matches 5586. .6316 of consensus"
repeat_region 29962. .30266
repeat_region /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region 30404. .30684
repeat_region /note="AluJo repeat: matches 5. .292 of consensus"
repeat_region 31091. .31221
repeat_region /note="L2 repeat: matches 2608. .2733 of consensus"
repeat_region 31721. .32315
repeat_region /note="L2 repeat: matches 1963. .2550 of consensus"
repeat_region 33041. .33560
repeat_region /note="MER67A repeat: matches 1. .543 of consensus"
repeat_region 34302. .35835
repeat_region /note="L1ME2 repeat: matches 4615. .6164 of consensus"
repeat_region 36046. .42211
repeat_region /note="L1P4 repeat: matches 1. .6144 of consensus"
misc_feature 42403. .42971
repeat_region /note="match: GSS: Em:AQ752221"
repeat_region 42913. .43588
repeat_region /note="L2 repeat: matches 2103. .2749 of consensus"
repeat_region 43631. .43861
repeat_region /note="L2 repeat: matches 1755. .1999 of consensus"
repeat_region 44481. .44588
repeat_region /note="L2 repeat: matches 2596. .2709 of consensus"
repeat_region 44598. .44728
repeat_region /note="MERSB repeat: matches 2. .134 of consensus"
repeat_region 44740. .45050
repeat_region /note="L1MC5 repeat: matches 7619. .7910 of consensus"
repeat_region 45051. .45350
repeat_region /note="AluSx repeat: matches 3. .303 of consensus"
repeat_region 45351. .45506
repeat_region /note="L1MC5 repeat: matches 7459. .7619 of consensus"
repeat_region 45507. .45532
repeat_region /note="13 copies 2 mer tt 100% conserved"
repeat_region 45543. .45716
repeat_region /note="Trig5 repeat: matches 889. .1063 of consensus"
repeat_region 45783. .46049
repeat_region /note="Trig5 repeat: matches 609. .881 of consensus"
repeat_region 46050. .46353
repeat_region /note="AluSc repeat: matches 1. .304 of consensus"
repeat_region 46354. .46989
repeat_region /note="Trig5 repeat: matches 1. .609 of consensus"
repeat_region 47003. .47196
repeat_region /note="L1MC/D repeat: matches 5258. .5453 of consensus"
repeat_region 47165. .47345
repeat_region /note="L1MC5 repeat: matches 7032. .7208 of consensus"
repeat_region 47791. .48163
repeat_region /note="MSTD repeat: matches 2. .390 of consensus"
repeat_region 48799. .48863
repeat_region /note="L1P3 repeat: matches 5687. .5751 of consensus"
repeat_region 48992. .49052
repeat_region /note="MERSB repeat: matches 2. .62 of consensus"
repeat_region 49893. .50028
```

Query Match 95.2%; Score 477; DB 9; Length 94727;  
Best Local Similarity 97.0%; Pred No. 3 6e-106;  
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 1 CTTTGGTGTTTACATGAAGTCCTTGGCCATGCTCTGATGCTGATGCTGCTG 60



```
Db 37088 CTTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTA 37147
QY 61 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTAAAGCTTTTAATCCATCTTG 120
Db 37148 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTAAAGCTTTTAATCCATCTTG 37207
QY 121 AATTAATTTTGTATAAGGTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGGCTA 180
Db 37208 AATTAATTTTGTATAAGGTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGGCTA 37267
QY 181 GCCAGTTTTCACAGACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTCT 240
Db 37268 GCCAGTTTTCACAGACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTCT 37327
QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTCGAGGCTCTGTC 300
Db 37328 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTCGAGGCTCTGTC 37387
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGCTAG 360
Db 37388 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGCTAG 37447
QY 361 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
Db 37448 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 37507
QY 421 GGATTGACTGCGCATGPGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTTT 480
Db 37508 GGATTGACTGCGCATGPGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTTT 37567
QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 37568 CCAATTCGTGAAGAAAGTAA 53588

RESULT 8
AC044782/c
LOCUS AC044782 149041 bp DNA linear PRI 23-APR-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-170M17, complete sequence.
ACCESSION AC044782
VERSION AC044782.6 GI:19774346
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 149041)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149041)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 149041)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 149041)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 5 (bases 1 to 149041)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Mar 26, 2002 this sequence version replaced gi:15004887.
```

```
FEATURES
source
Location/Qualifiers
1..149041
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-170M17"
/clone_lib="RPC1-11"
BASE COUNT 46956 a 31551 c 29426 g 41108 t
ORIGIN

Query Match 95.28; Score 477; DB 9; Length 149041;
Best Local Similarity 97.08; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CTTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTG 60
Db 53820 CTTTTGGTGTATAGACATGAAGTCCCTTGCCCATGCCTATGCTCCTGAATGCTATGGCTA 53761
QY 61 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTTAAAGCTTTTAATCCATCTTG 120
Db 53760 GGTTCCTCTAGGAGATTTATGTTTATGCTCTAAGCTTTAAAGCTTTTAATCCATCTTG 53701
QY 121 AATTAATTTTGTATAAGGTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGGCTA 180
Db 53700 AATTAATTTTGTATAAGGTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGGCTA 53641
QY 181 GCCAGTTTTCACAGACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTCT 240
Db 53640 GCCAGTTTTCACAGACCATTTATTAATAGGGAATCCTTCCCATGCTGTTTCT 53581
QY 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTCGAGGCTCTGTC 300
Db 53580 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCATTTATTCGAGGCTCTGTC 53521
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGCTAG 360
Db 53520 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGCTAG 53461
QY 361 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
Db 53460 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGCCCTCCAGCTTGTCTTTGGCTTA 53401
QY 421 GGATTGACTGCGCATGPGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTTT 480
Db 53400 GGATTGACTGCGCATGPGGCTCTTTTGGTCCACATCAACTTTAAAGTACTGTTTT 53341
QY 481 CCAATTCGTGAAGAAAGTAA 501
Db 53340 CCAATTCGTGAAGAAAGTAA 53320
```

```
RESULT 9
AC023273/c
LOCUS AC023273 154076 bp DNA linear HTG 21-FEB-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-199D16, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION AC023273
VERSION AC023273.6 GI:13027570
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 154076)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154076)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```

COMMENT

On Feb 21, 2001 this sequence version replaced g1:11527490.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0199D16  
----- Summary Statistics -----

Sequencing vector: M13, 85%  
Sequencing vector: plasmid, 8%  
Chemistry: Dye-terminator ET; 85% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 138250 bases at least Q40  
Consensus quality: 143237 bases at least Q30  
Consensus quality: 146644 bases at least Q20  
Insert size: 142000; agarose-fp  
Insert size: 152276; sum-of-contigs  
Quality coverage: 4.28 in Q20 bases; agarose-fp  
Quality coverage: 3.94 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1706: contig of 1706 bp in length  
1707 1806: gap of unknown length  
1807 3314: contig of 1508 bp in length  
3315 3415: gap of unknown length  
3416 5225: contig of 1811 bp in length  
5226 5325: gap of unknown length  
5326 6708: contig of 1382 bp in length  
6709 6808: gap of unknown length  
6809 8647: contig of 1840 bp in length  
8648 8747: gap of unknown length  
8748 11645: contig of 2898 bp in length  
11646 11745: gap of unknown length  
11746 16416: contig of 4671 bp in length  
16417 16517: gap of unknown length  
16518 22032: contig of 5516 bp in length  
22033 22133: gap of unknown length  
22134 26662: contig of 4530 bp in length  
26663 26763: gap of unknown length  
26764 31641: contig of 4879 bp in length  
31642 31741: gap of unknown length  
31742 37785: contig of 6043 bp in length  
37786 37885: gap of unknown length  
37886 46462: contig of 8577 bp in length  
46463 46562: gap of unknown length  
46563 56411: contig of 9849 bp in length  
56412 56511: gap of unknown length  
56512 68341: contig of 11831 bp in length  
68342 68441: gap of unknown length  
68442 82114: contig of 13673 bp in length  
82115 82215: gap of unknown length  
82216 96726: contig of 14512 bp in length  
96727 96826: gap of unknown length  
96827 113874: contig of 17048 bp in length  
113875 113975: gap of unknown length  
113976 131413: contig of 17438 bp in length  
131414 131513: gap of unknown length  
131514 154076: contig of 22564 bp in length.  
Location/Qualifiers  
1. 154076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-199D16"

FEATURES  
source

misc\_feature 1. 1706  
/note="assembly\_name:Contig33"  
misc\_feature 1807. 3314  
/note="assembly\_name:Contig34"  
misc\_feature 3415. 5225  
/note="assembly\_name:Contig35"  
misc\_feature 5326. 6707  
/note="assembly\_name:Contig36"  
misc\_feature 6808. 8647  
/note="assembly\_name:Contig37"  
misc\_feature 8748. 11645  
/note="assembly\_name:Contig38"  
misc\_feature 11746. 16416  
/note="assembly\_name:Contig39  
clone\_end:SP6  
vector\_side:left  
16517. 22032  
/note="assembly\_name:Contig40"  
misc\_feature 22133. 26662  
/note="assembly\_name:Contig41"  
misc\_feature 26763. 31641  
/note="assembly\_name:Contig42"  
misc\_feature 31742. 37784  
/note="assembly\_name:Contig43"  
misc\_feature 37885. 46461  
/note="assembly\_name:Contig44"  
misc\_feature 46562. 56410  
/note="assembly\_name:Contig45"  
misc\_feature 56511. 68341  
/note="assembly\_name:Contig46"  
misc\_feature 68442. 82114  
/note="assembly\_name:Contig47"  
misc\_feature 82215. 96726  
/note="assembly\_name:Contig48"  
misc\_feature 96827. 113874  
/note="assembly\_name:Contig49"  
misc\_feature 113975. 131412  
/note="assembly\_name:Contig50"  
misc\_feature 131513. 154076  
/note="assembly\_name:Contig51"  
BASE COUNT 44022 a 33471 c 33238 g 41531 t 1814 others  
ORIGIN

Query Match 95.2%; Score 477; DB 2; Length 154076;

Best Local Similarity 97.0%; Pred. No. 3.4e-106;

Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTCTGTAATGTTGCTG 60  
|||||  
Db 9271 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTCTGTAATGTTGCTG 9212  
Qy 61 GGTTCCTTCTAGGATTTTATGTTTGTAGTCTAAGCTTTAAGCTTTAATCCATCTG 120  
|||||  
Db 9211 GGTTCCTTCTAGGATTTTATGTTTGTAGTCTAAGCTTTAAGCTTTAATCCATCTG 9152  
Qy 121 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180  
|||||  
Db 9151 AATTAATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 9092  
Qy 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTGTTTCT 240  
|||||  
Db 9091 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATGCTGTTTCTG 9032  
Qy 241 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATATTTCTCAGGGCTCTGTC 300  
|||||  
Db 9031 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATATTTCTCAGGGCTCTGTC 8972  
Qy 301 TGTTCATTTGTTGATATCTCTGTTTGTGGTACCAAGTACCATGTTGTTTGGTACTGTAG 360  
|||||  
Db 8971 TGTTCATTTGTTGATATCTCTGTTTGTGGTACCAAGTACCATGTTGTTTGGTACTGTAG 8912  
Qy 361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGTCCTCCAGCTTTCTCTTTGGCTTA 420  
|||||

```

Db      8911 CTTTGTAGTATAGTTTGAAGTCAGGTAGCGTCCAGCTTCAGCTTCTCTTTGGCTTA 8852
Qy      421 GGATTGACTTGGCGATCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      8851 GGACTGACTTGGCGATCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 8792
Qy      481 CCAATTCCTGTGAAGAAAGTAA 501
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      8791 CCAATTCCTGTGAAGAAAGTCA 8771

RESULT 10
AC099504/c
LOCUS      162841 bp      DNA      linear      PRI 15-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-219L6, complete sequence.
AC099504
AC099504.2 GI:18677371
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 162841)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Unpublished
            2 (bases 1 to 162841)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 162841)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (15-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Feb 15, 2002 this sequence version replaced gi:16930920.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE: This insert is not the entire sequence of the clone (entire
            sequence is 173kb). It is clipped at the overlap with AC008558. The
            number of bases overlapped is 33072.

FEATURES             source
            Location/Qualifiers
            1..162841
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="5"
               /clone="RP11-219L6"
BASE COUNT  52128 a 29187 c 29340 g 52186 t
ORIGIN
Query Match      95.2%; Score 477; DB 9; Length 162841;
Best Local Similarity 97.0%; Pred. No. 3.4e-106;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy      1 CTTTGGTGTATTAGACATGAAGTCCTGTCCTGATGCTGCTGAAGGTATGCTG 60
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32925 CTTTGGTGTATTAGACATGAAGTCCTGTCCTGATGCTGCTGAAGGTATGCTG 32866
Qy      61 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32865 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 32806
Qy      121 AATTAATTTTGTATAGGTGTGAAGGATCCAGTTTTCAGCTTTTCATACATAGGGCTA 180
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32805 AATTAATTTTGTATAGGTGTGAAGGATCCAGTTTTCATACATAGGGCTA 32746
Qy      181 GCACGTTTTCTACGACCATTTATTAAGTAGGAATCCCTTCCCAATGCTGTTGTTTTCT 240
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII

```

```

Db      32745 GCACGTTTTCCAGCACCAATTTATTAATAGGGAATCCTTCCCAATGCTGTTTCT 32686
Qy      241 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGCGCATATTTCTTGAGGCTCTGTTTC 300
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32685 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGCGCATATTTCTTGAGGCTCTGTTTC 32626
Qy      301 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTATGTTTGGTGTACTGTAG 360
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32625 CGTTCCATTTGGTCTATATATCTCTGTTTGGTACCAGTACCAGTATGTTTGGTGTACTGTAG 32566
Qy      361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTCTTTGGGCTTA 420
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32565 CCTTGTAGTATAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTCTTTGGGCTTA 32506
Qy      421 GGATTGACTTGGCGATGCTGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32505 GGATTGACTTGGTGTATGCTGGCTCTTTTGGTTCATATGAACCTTTAAAGTAGTTTTT 32446
Qy      481 CCAATTCCTGTGAAGAAAGTAA 501
      IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db      32445 CCAATTCCTGTGAAGAAAGTCA 32425

RESULT 11
AC019053/c
LOCUS      169161 bp      DNA      linear      HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-96J19, WORKING DRAFT SEQUENCE,
            13 unordered pieces.
AC019053
AC019053.3 GI:8348015
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 169161)
            Waterston, R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 169161)
            Waterston, R.H.
            Direct Submission
            Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Jun 8, 2000 this sequence version replaced gi:7230878.

COMMENT
            ----- Genome Center -----
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----- Project Information -----
            Center project name: H_NH0096J19
            ----- Summary Statistics -----
            Sequencing vector: M13; 86%
            Chemistry: Dye-primer ET; 86% of reads
            Chemistry: Dye-terminator Big Dye; 14% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 158339 bases at least Q40
            Consensus quality: 162156 bases at least Q30
            Consensus quality: 164609 bases at least Q20
            Insert size: 166000; agarose-fp
            Insert size: 167961; sum-of-contigs
            Quality coverage: 4.70 in Q20 bases; agarose-fp
            Quality coverage: 4.11 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.

1221: contig of 1221 bp in length  
1321: gap of unknown length  
5345: contig of 4024 bp in length  
5445: gap of unknown length  
9464: contig of 4019 bp in length  
9564: gap of unknown length  
12934: contig of 3370 bp in length  
13034: gap of unknown length  
12935: contig of 9514 bp in length  
22848: gap of unknown length  
22849: contig of 11388 bp in length  
34036: gap of unknown length  
34037: contig of 11094 bp in length  
45230: gap of unknown length  
45231: contig of 11998 bp in length  
57328: gap of unknown length  
57428: gap of unknown length  
76592: contig of 19164 bp in length  
76593: gap of unknown length  
96129: contig of 19437 bp in length  
96229: gap of unknown length  
115902: contig of 19673 bp in length  
115903: gap of unknown length  
116002: gap of unknown length  
136908: contig of 20906 bp in length  
137008: gap of unknown length  
137009: 169161: contig of 32153 bp in length.

## FEATURES

Source

1. 169161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-96J19"  
1. 1221  
/note="assembly\_name:Contig7  
clone\_end:T7  
vector\_side:right"

misc\_feature

1322..5345  
/note="assembly\_name:Contig8"  
5446..9464  
/note="assembly\_name:Contig9"  
9565..12934  
/note="assembly\_name:Contig11"  
13035..22548  
/note="assembly\_name:Contig12"  
22649..34036  
/note="assembly\_name:Contig13  
clone\_end:SP6  
vector\_side:right"  
34137..45230  
/note="assembly\_name:Contig14"  
45331..57328  
/note="assembly\_name:Contig15"  
57428..76592  
/note="assembly\_name:Contig16"  
76693..96129  
/note="assembly\_name:Contig17"  
96230..115902  
/note="assembly\_name:Contig18"  
116003..136908  
/note="assembly\_name:Contig19"  
137009..169161  
/note="assembly\_name:Contig20"

BASE COUNT 53688 a 34273 c 33376 g 47619 t 1205 others  
ORIGIN

Query Match 95.2%; Score 477; DB 2; Length 169161;  
Best Local Similarity 97.0%; Pred. No. 3.4e-106;  
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
1 CTTTGGGCTTTACACATGAAGTCTTCCCATGCTATGCTGCTGAATGGTATGCGG 60  
|||||  
96067 CTTTGGGCTTTACACATGAAGTCTTCCCATGCTATGCTGCTGAATGGTATGCGG 60  
|||||

QY

DB

QY	61	GGTTTCTCTAGGATTTTATGGTCTAACATTTAAGTCTTTAAATCATCTTG	120
DB	96007	GGTTTCTCTAGGATTTTATGGTCTAACATTTAAGTCTTTAAATCATCTTG	95948
QY	121	AATTAATTTTGTATAAGGTGAAGAGGGATCCAGTTTCAGCTTTCACATAGGCGTA	180
DB	95947	AATTAATTTTGTATAAGGTGAAGAGGGATCCAGTTTCAGCTTTCACATAGGCGTA	95888
QY	181	GCCAGTTTCTCAGCACCATTATTAAATAGGAATCCCTCCCATTCCTGTTTTTCT	240
DB	95887	GCCAGTTTCTCAGCACCATTATTAAATAGGAATCCCTCCCATTCCTGTTTTTCT	95828
QY	241	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGCGCATATTATTCAGGGGCTCTGTT	300
DB	95827	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGCGCATATTATTCAGGGGCTCTGTT	95768
QY	301	TGTTCCATTTGGTGTATATCTCTGTTTGTACACAGTACCAGTGTGTTTGGTTACTGTAG	360
DB	95767	TGTTCCATTTGGTGTATATCTCTGTTTGTACACAGTACCAGTGTGTTTGGTTACTGTAG	95708
QY	361	CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTCTTTTGGCTTA	420
DB	95707	CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTCTTTTGGCTTA	95648
QY	421	GGATTGACCTTGGCGATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT	480
DB	95647	GGATTGACCTTGGCGATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT	95588
QY	481	CCAAATCTCTGAAGAAAGTAA 501	
DB	95587	CCAAATCTCTGAAGAAAGTCA 95567	

RESULT 12  
AC068209/c  
LOCUS  
DEFINITION Homo sapiens chromosome 2 clone RP11-25018 map 2, WORKING DRAFT  
SEQUENCE, 25 unordered pieces.  
AC068209  
AC068209.3 GI:10280752  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 171490)  
Birren.B., Linton.L., Nusbaum.C. and Lander.E.  
Homo sapiens chromosome 2, clone RP11-25018  
Unpublished  
REFERENCE 2 (bases 1 to 171490)  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,  
Anderson.S., Baldwin.J., Barna.N., Bastien.V., Beda.F.,  
Boguslavsky.L., Boukhgalter.B., Brown.A., Burkett.G.,  
Campopiano.A., Castle.A., Choepe.Y., Colangelo.M., Collins.S.,  
Collymore.A., Cooke.P., DeArrellano.K., Dewar.K., Diaz.J.S.,  
Dodge.S., Domino.M., Doyle.M., Ferreira.P., FitzHugh.W., Gage.D.,  
Galagan.J., Gardyna.S., Ginde.S., Goyette.M., Graham.L.,  
Grand-Pierre.N., Grant.G., Hagos.B., Heaford.A., Horton.L.,  
Howland.J.C., Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A.,  
Klein.J., LaRocque.K., Lamazares.R., Landers.T., Lehoczy.J.,  
Levine.R., Liu.C., Liu.G., Locke.K., Macdonald.P., Marquis.N.,  
McCarthy.M., McEwan.P., McGurk.A., McKernan.K., McPheeters.R.,  
Meldrum.J., Meneus.L., Mihova.T., Miranda.C., Mieng.V., Morrow.J.,  
Murphy.T., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,  
O'Neill.D., Oliver.T.M., Oliver.J., Peterson.K., Pierre.N.,  
Pisani.C., Pollara.V., Raymond.C., Riley.R., Rogov.P., Rothman.D.,  
Roy.A., Santos.R., Schauer.S., Severy.P., Spencer.B.,  
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,  
Tesfaye.S., Theodore.J., Tirrell.A., Travers.M., Trigilio.J.,  
Vassiliev.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,  
Young.G., Zainoun.J., Zimmer.A. and Zody.M.  
Direct Submission

TITLE

**JOURNAL**

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 23, 2000 this sequence version replaced gi:85669120.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

**COMMENT**

On Sep 23, 2000 this sequence version replaced gi:8569120.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10069

```
Center clone name: 250-I_8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: dye-terminator Big dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157936 bases at least Q40
Consensus quality: 165044 bases at least Q30
Consensus quality: 167707 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 169090; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
```

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. As soon as it is available and the accession number will be preserved.

*	2797	2896:	gap of	2796:	contig of	2796 bp	in length
*	2897	2896:	gap of	100 bp			
*	2897	4248:	contig of	1352 bp	in length		
*	4249	4348:	gap of	100 bp			
*	4349	5349:	contig of	1001 bp	in length		
*	5350	5449:	gap of	100 bp			
*	5450	7045:	contig of	1596 bp	in length		
*	7046	7145:	gap of	100 bp			
*	7146	8600:	contig of	1455 bp	in length		
*	8601	8700:	gap of	100 bp			
*	8701	11899:	contig of	3199 bp	in length		
*	11900	11999:	gap of	100 bp			
*	12000	15340:	contig of	3341 bp	in length		
*	15341	15440:	gap of	100 bp			
*	15441	17756:	contig of	2316 bp	in length		
*	17757	17856:	gap of	100 bp			
*	17857	22299:	contig of	4443 bp	in length		
*	22300	22399:	gap of	100 bp			
*	22400	25462:	contig of	3063 bp	in length		
*	25463	25562:	gap of	100 bp			
*	25563	28942:	contig of	3380 bp	in length		
*	28943	29042:	gap of	100 bp			
*	29043	34484:	contig of	5442 bp	in length		
*	34485	34584:	gap of	100 bp			
*	34585	38797:	contig of	4213 bp	in length		
*	38798	38897:	gap of	100 bp			
*	38898	43155:	contig of	6258 bp	in length		
*	43156	45255:	gap of	100 bp			
*	45256	50138:	contig of	4883 bp	in length		
*	50139	50238:	gap of	100 bp			
*	50239	57081:	contig of	6843 bp	in length		
*	57082	57181:	gap of	100 bp			
*	57182	62192:	contig of	5011 bp	in length		
*	62193	62292:	gap of	100 bp			
*	62293	68432:	contig of	6140 bp	in length		
*	68433	68532:	gap of	100 bp			
*	68533	78869:	contig of	10337 bp	in length		
*	78870	78969:	gap of	100 bp			
*	78970	85886:	contig of	6917 bp	in length		

BASE COUNT	50305 a	35432 c	33845 g	49505 t	2403 others
vectorside: 119mc					

22



Query Match 95.2%; Score 477; DB 2; Length 176630;  
Best Local Similarity 97.0%; Pred. No. 3.4e-106;  
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CTTTGGTGTAGACATGAAGTCCCTTGGCCATGCTATGCTGCTGAATGCTATGCTG 60  
DB CTTTGGTGTGTAGACATGAAGTCCCTTGGCCATGCTATGCTGCTGAATGCTATGCTG 7151  
QY 61 GGTTCCTCTCAGGAGATTTATGCTTAAAGTCTAAAGTCTTAAAGTCTTAAAGTCTT 120  
DB GGTTCCTCTCAGGAGATTTATGCTTAAAGTCTAAAGTCTTAAAGTCTTAAAGTCTT 7091  
QY 121 AATTAATTTTGTATAAGGTCTAAGGAAGGATCCAGTTTCTACATGAGGCTA 180  
DB AATTAATTTTGTATAAGGTCTAAGGAAGGATCCAGTTTCTACATGAGGCTA 7031  
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTCCCATGCTGTTTCT 240  
DB GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTCCCATGCTGTTTCT 6971  
QY 241 CAGGTTTCTCAAGATCAGATAGTGTAGATATGCGGATATTTCTGAGGGCTCTGTT 300  
DB CAGGTTTCTCAAGATCAGATAGTGTAGATATGCGGATATTTCTGAGGGCTCTGTT 6911  
QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360  
DB TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 6851  
QY 361 CTTTGTAGTGTAGTTGAAGTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
DB CTTTGTAGTGTAGTTGAAGTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 6791  
QY 421 GGATGACTTGGGATGTTGGGCTCTTTTGGTCCACATGAAGTCTTAAAGTACTTTT 480  
DB GGATGACTTGGGATGTTGGGCTCTTTTGGTCCACATGAAGTCTTAAAGTACTTTT 6731  
QY 481 CCAATTCGTGAAGAAAGTAA 501  
DB CCAATTCGTGAAGAAAGTCA 6710

RESULT 14  
AC069543 179789 bp DNA linear PRI 28-MAR-2002  
LOCUS Homo sapiens chromosome 10 clone RP11-393H5, complete sequence.  
DEFINITION AC069543  
ACCESSION AC069543.5 GI:19774362  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 179789)  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 179789)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
REFERENCE 3 (bases 1 to 179789)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA  
REFERENCE 4 (bases 1 to 179789)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

COMMENT On Mar 28, 2002 this sequence version replaced gi:14010785.  
FEATURES  
source 1. 179789  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-393H5"  
/clone\_lib="RPC1-11"  
BASE COUNT 50198 a 34202 c 37287 g 58102 t  
ORIGIN

Query Match 95.2%; Score 477; DB 9; Length 179789;  
Best Local Similarity 97.0%; Pred. No. 3.4e-106;  
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTAGACATGAAGTCCCTTGGCCATGCTATGCTGCTGAATGCTATGCTG 60  
DB 153506 CTTTGGTGTGTAGACATGAAGTCCCTTGGCCATGCTATGCTGCTGAATGCTATGCTG 153565  
QY 61 GGTTCCTCTCAGGAGATTTATGCTTAAAGTCTAAGATTTAAGTCTTAAAGTCTTAA 120  
DB 153566 GGTTCCTCTCAGGAGATTTATGCTTAAAGTCTAAGATTTAAGTCTTAAAGTCTTAA 153625  
QY 121 AATTAATTTTGTATAAGGTCTAAGGAAGGATCCAGTTTCTACATGAGGCTA 180  
DB 153626 AATTAATTTTGTATAAGGTCTAAGGAAGGATCCAGTTTCTACATGAGGCTA 153685  
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTCCCATGCTGTTTCT 240  
DB 153686 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTCCCATGCTGTTTCT 153745  
QY 241 CAGGTTTGTCAAGATCAGATAGTGTAGATATGCGGATATTTCTGAGGGCTCTGTT 300  
DB 153746 CAGGTTTGTCAAGATCAGATAGTGTAGATATGCGGATATTTCTGAGGGCTCTGTT 153805  
QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360  
DB 153806 TGTTCATTTGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 153865  
QY 361 CTTTGTAGTGTAGTTGAAGTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
DB 153866 CTTTGTAGTGTAGTTGAAGTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCT 153925  
QY 421 GGATGACTTGGGATGTTGGGCTCTTTTGGTCCACATGAAGTCTTAAAGTACTTTT 480  
DB 153926 GGATGACTTGGGATGTTGGGCTCTTTTGGTCCACATGAAGTCTTAAAGTACTTTT 153985  
QY 481 CCAATTCGTGAAGAAAGTAA 501  
DB 153986 CCAATTCGTGAAGAAAGTCA 154006

RESULT 15  
AC087774 184000 bp DNA linear PRI 07-MAR-2002  
LOCUS Homo sapiens 12p BAC RP11-633013 (Roswell Park Cancer Institute  
DEFINITION Human BAC Library) complete sequence.  
ACCESSION AC087774  
VERSION AC087774.27 GI:16572889  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 184000)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,



Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flag, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oraguine, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zucherlapati, R., Weinstein, G. and Gibbs, R.

Direct Submission  
Unpublished to 184000)

Worley, K.C.  
Direct Submission  
Submitted (24-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 184000)  
Worley, K.C.  
Direct Submission  
Submitted (01-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 184000)  
Worley, K.C.  
Direct Submission  
Submitted (02-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 184000)  
Worley, K.C.  
Direct Submission  
Submitted (26-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 184000)  
Worley, K.C.  
Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 1, 2001 this sequence version replaced gi:16327961.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [9c-help@bcm.tmc.edu](mailto:9c-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
		1. .184000
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosomes="12"
		/clone="RP11-633013"
misc_feature		1. .2009
		/notes="overlaps bases 18791. .20794 of clone AC093950"
		/function="clone overlap"
repeat_region		26. .46
		/rpt_family="AT_rich"
repeat_region		1466. .1804
		/rpt_family="LIP3"
repeat_region		2274. .2583
		/rpt_family="AluSp"
misc_feature		2584
		/notes="(A or *)"
		/function="unsure base"
repeat_region		complement(2851. .2921)
		/rpt_family="LIP"
repeat_region		2922. .5243
		/rpt_family="LIP3"
repeat_region		complement(5470. .6019)
		/rpt_family="MLTID"
repeat_region		6640. .6756
		/rpt_family="HERVL"
repeat_region		7042. .7074
		/rpt_family="(TTTA)n"
repeat_region		complement(7092. .7265)
		/rpt_family="LIHS"
repeat_region		7563. .8070
		/rpt_family="LTR40a"
repeat_region		8174. .8253
		/rpt_family="MERVL"
repeat_region		8254. .8315
		/rpt_family="(TA)n"
repeat_region		8316. .8541
		/rpt_family="MERVL"
repeat_region		8791. .9287
		/rpt_family="MLTIC"
repeat_region		9497. .9538
		/rpt_family="AT_rich"

```

Db          71085 CCAATTCTGTGAAGAAAGTCA 71105
|||||
Search completed: July  4, 2003, 16:40:47
Job time : 1513.67 secs

```

repeat_region	9782. .9982		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(10662. .10772)		
repeat_region	/rpt_family="MIR"		
repeat_region	11333. .11367		
misc_feature	/rpt_family="AT-rich"		
misc_feature	11368		
misc_feature	/note="(A or *)"		
repeat_region	/function="unsure base"		
repeat_region	complement(11808. .12111)		
repeat_region	/rpt_family="AluY"		
repeat_region	12258. .12424		
repeat_region	/rpt_family="(TA)n"		
repeat_region	complement(12617. .12916)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(13063. .13267)		
repeat_region	/rpt_family="MER58A"		
repeat_region	complement(13271. .13711)		
repeat_region	/rpt_family="LIME1"		
repeat_region	complement(13764. .13911)		
repeat_region	/rpt_family="L1"		
repeat_region	14246. .14287		
repeat_region	/rpt_family="AT-rich"		
repeat_region	14862. .14883		
repeat_region	/rpt_family="AT-rich"		
repeat_region	complement(15988. .16201)		
repeat_region	/rpt_family="MIR"		
repeat_region	17159. .17448		
repeat_region	/rpt_family="LIMA4"		
repeat_region	17910. .17932		
repeat_region	/rpt_family="AT-rich"		
repeat_region	complement(17933. .19771)		
repeat_region	/rpt_family="LIMA4"		
repeat_region	21371. .21519		
repeat_region	/rpt_family="Lip"		
Query Match 95.2%; Score 477; DB 9; Length 184000;			
Best Local Similarity 97.0%; Pred. No. 3.4e-106;			
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;			
Qy	1	CTTTTGGTGTTTAGACATGAAGTCCTTCCCATGCCTATGTCCTGAATGGTATTCCTG 60	
Db	70605	CTTTTGGTGTTTAGACATGAAGTCCTTCCCATGCCTATGTCCTGAATGGTATTCCTA 70604	
Qy	61	GGTTTCTCTTAGGGATTTATGGTTTATAGGCTAACATTTAAAGCTTTTATCCATCTTG 120	
Db	70665	GGTTTCTCTTAGGGATTTATGGTTTATAGGCTAACATTTAAAGCTTTTATCCATCTTG 70724	
Qy	121	AATTAATTTTGTATAGGTGAAGGATCCAGTTTCCAGCTTTCTACATAGGGCTA 180	
Db	70725	AATTGATTTTGTATAGGTGAAGGATCCAGTTTCCAGCTTTCTACATAGGGCTA 70784	
Qy	181	GCAGTTTTCTCAGCACCAATTTATTAATAGGAATCCCTTTCCCATTCGTTGTTTTCT 240	
Db	70785	GCAGTTTTCTCAGCACCAATTTATTAATAGGAATCCCTTTCCCATTCGTTGTTTTCT 70844	
Qy	241	CAGTTTTGTCAAGATCAGATAGTGTAGATATGCGGATATTTCTCAGGCGCTCTGTC 300	
Db	70845	CAGTTTTGTCAAGATCAGATAGTGTAGATATGCGGATATTTCTCAGGCGCTCTGTC 70904	
Qy	301	TGTTCCATTTGGTTGATATCTCTGTTTTGGTACCAGTACCATGTTGTTGGTTACTGTAG 360	
Db	70905	TGTTCCATTTGATATATCTCTGTTTTGGTACCAGTACCATGTTGTTGGTTACTGTAG 70964	
Qy	361	CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTGTTGGCTTA 420	
Db	70965	CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTGTTGGCTTA 71024	
Qy	421	GGATTGACATGGGATGTCGGCTCTTTTGGTCCACATGACATGACATTTAAAGTACTTTT 480	
Db	71025	GGATTGACATGGGATGTCGGCTCTTTTGGTCCATGATGACATTTAAAGTACTTTT 71084	
Qy	481	CCAATTCCTGTGAAGAAGTAA 501	

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 14, 2003, 15:09:32 ; Search time 1093.54 seconds  
(without alignments)  
7405.060 Million cell updates/sec

Title: US-10-083-853b-2\_COPY\_1\_500

Perfect score: 500

Sequence: 1 gtagatgaagaagcctca.....actaaagttcttcaggct 500

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST

- 1: em\_estb1.\*
- 2: em\_estb1.\*
- 3: em\_estb1.\*
- 4: em\_estb1.\*
- 5: em\_estb1.\*
- 6: em\_estb1.\*
- 7: em\_estb1.\*
- 8: em\_estb1.\*
- 9: gblest1.\*
- 10: gb\_est2.\*
- 11: gb\_est2.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estfun.\*
- 17: gb\_gss.\*
- 18: em\_gss\_hum.\*
- 19: em\_gss\_inv.\*
- 20: em\_gss\_pln.\*
- 21: em\_gss\_vrt.\*
- 22: em\_gss\_fun.\*
- 23: em\_gss\_mam.\*
- 24: em\_gss\_mus.\*
- 25: em\_gss\_othr.\*
- 26: em\_gss\_pro.\*
- 27: em\_gss\_rtd.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	500	100.0	606	14	BQ575044	BQ575044 UI-H-E21-
2	500	100.0	3739	11	AF113216	AF113216 Homo sapi
3	471.8	94.4	505	10	AW274723	AW274723 xn34b02.x
4	460	92.0	485	9	AA488996	AA488996 aa54c10.x
5	370.8	74.2	406	9	AI221902	AI221902 qg99h07.x
6	250.8	50.2	467	13	BM144799	BM144799 TCAAP1D13

C	7	244.4	48.9	560	10	AW968905
8	202.4	40.5	217	9	AI570720	
9	154.6	30.9	318	9	AA904948	
10	149.4	29.9	317	9	AA736766	
11	131.8	26.4	360	9	AA488780	
C	12	108.6	21.7	1796	11	AK003359
13	69.4	13.9	924	17	CNS0206S	
C	14	66.6	13.3	1124	17	CNS073BM
C	15	64.4	12.9	1201	17	CNS016E1
C	16	63.2	12.6	1043	17	CNS0145P
C	17	62.6	12.5	1101	17	CNS00EVL
C	18	62	12.4	945	17	CNS04D0K
19	60	12.0	1084	17	CNS071NH	
C	20	60	12.0	1200	17	CNS016CO
C	21	57.6	11.5	450	13	BI297936
22	57.6	11.5	945	17	CNS04D0K	
C	23	57.2	11.4	1101	17	CNS0039G
C	24	57	11.4	1101	17	CNS00E07
C	25	56.6	11.3	1092	17	CNS020K7
26	56.4	11.3	928	17	CNS00DKY	
C	27	56.4	11.3	1101	17	CNS0039G
C	28	56.2	11.2	418	13	BI297322
29	56.2	11.2	427	13	BI297805	
C	30	56.2	11.2	466	13	BI297955
C	31	56.2	11.2	482	12	BF406722
C	32	56.2	11.2	489	9	AA997136
C	33	56.2	11.2	566	12	BG374585
C	34	56.2	11.2	571	13	BI301086
C	35	56.2	11.2	550	13	BI296665
C	36	56.2	11.2	758	13	BI296653
37	55.6	11.1	1101	17	CNS0172Q	
38	54.8	11.0	676	17	BH648307	
C	39	54.6	10.9	428	13	BI297921
C	40	54.6	10.9	441	12	BG374397
C	41	54.4	10.9	1037	14	BQ648574
42	54	10.8	1201	17	CNS016E1	
C	43	53.8	10.8	1013	17	CNS06RPQ
C	44	53.6	10.7	928	17	CNS00DKY
C	45	53.4	10.7	1101	17	CNS00EVL

ALIGNMENTS

RESULT 1  
BQ575044  
LOCUS  
DEFINITION  
UI-H-E21-bbd-o-15-0-UI-s1 NCI-CGAP-Ch2 Homo sapiens cDNA clone  
UI-H-E21-bbd-o-15-0-UI 3', mRNA sequence.  
ACCESSION  
BQ575044  
VERSION  
BQ575044.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 606)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov).  
Tissue procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of  
Orthopaedics  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following repetitive elements were found in this cDNA  
sequence: 1-24, >AT-rich#Low-complexity 122-142,  
>AT-rich#Low-complexity 249-325, >(TA)n#Simple\_repeat

606 bp  
linear  
EST 19-JUN-2002



```

Db      1323 CAGGCTAGTAAGTTGGTGGTAGGAAAAGGCTCTCTTATCTCACCCCTCTTAA 1264
QY      481 ACTAAAGCTTTCTTCAGGCT 500
Db      1263 ACTAAAGCTTTCTTCAGGCT 1244

RESULT 3
LOCUS   AW274723
DEFINITION x34b02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2695563 3',
mRNA sequence.
ACCESSION AW274723.1
VERSION   1
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: 40UP from Gibco
High quality sequence stop: 475.
Location/Qualifiers
1. 505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2695563"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 132376-132391, 145607-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 169 a 83 c 75 g 178 t
ORIGIN
Query Match 94.4%; Score 471.8; DB 10; Length 505;
Best Local Similarity 99.2%; Pred. No. 6.4e-93;
Matches 485; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY      1 GTATATGTAAGAACGCCCTCACTCTTTTGATTTTAAATACAGATGCTTCTTTAAAGAGA 60
Db      17 GTATATGTAAGAACGCCCTCACTCTTTTGATTTTAAATACAGATGCTTCTTTAAAGAGA 76
QY      61 GCAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTCTAAATTTT 120
Db      77 GCAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATTAATTTATCTCTCTAAATTTT 136
QY      121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 180
Db      137 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGGCAAGGATTTTAAGAGTCT 196
QY      181 AACTCAACACATGATGTAAGCTCTGGTGTAACCTGGTATATATATACC--AAAAAACAATTTG 238

```

```

Db      137 AACTCAACACATGATGTAAGCTCTGGTGTAACCTGGTATATATATACCAAAAAAACAATTTG 256
QY      239 ATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAA 298
Db      257 ATCTATATACACATAGACATGAATATATTTCTGTGTGTGTGTGTGCATATATAACCTCAA 316
QY      299 ACACATATTATTAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTTCT 358
Db      317 ACACATATTATTAATGCAATCCTATATTTCTTAGGTATAGAAAGTTGATGATATACCTTTCT 376
QY      359 ACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGCAT 418
Db      377 ACTTGCCATGGCATTAAACAAAGCAAGGCTGAGACTCAGCAACCACTTGTGTTCAATGCAT 436
QY      419 TGCAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAGGTTCTCTATATCTACCCCTCTT 478
Db      437 TGCAGGCTAGTAGTAAGTTTGGTTCCTGGTAGGAAAAGGTTCTCTATCTACCCCTCTT 496
QY      479 AAACATAAAG 487
Db      497 AAACATAAAG 505

RESULT 4
LOCUS   AA488996
DEFINITION aa54c10.s1 NCI_CGAP_GCBI Homo sapiens cDNA clone IMAGE:824754 3',
mRNA sequence.
ACCESSION AA488996
VERSION   AA488996.1
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -4lm13 fwd. Et from Amersham
High quality sequence stop: 465.
Location/Qualifiers
1. 485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:824754"
/clone_lib="NCI_CGAP_GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library

```

went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 161 a 72 c 75 g 177 t

ORIGIN

Query Match 92.0%; Score 460; DB 9; Length 485;  
Best Local Similarity 100.0%; Pred. No. 2.4e-90;  
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTAAGAGA 60  
|||||  
Db 26 GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTAAGAGA 85  
|||||

QY 61 GCAAGATTCAAATGTTTGTCTTTCACAAATTTAAATATAATATATCTCTTAAATTTT 120  
|||||  
Db 86 GCAAGATTCAAATGTTTGTCTTTCACAAATTTAAATATAATATATCTCTTAAATTTT 145  
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180  
|||||  
Db 146 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 205  
|||||

QY 181 AACTCAAAACATATGTAAGCTCTGCTGTACCTGTTATATATACCAAAAAACATTTGAT 240  
|||||  
Db 206 AACTCAAAACATATGTAAGCTCTGCTGTACCTGTTATATATACCAAAAAACATTTGAT 265  
|||||

QY 241 CTATATACACATAGACATGAATATATTTCTGTGCTGTTTGTGCATATATAAACCCTCAAA 300  
|||||  
Db 266 CTATATACACATAGACATGAATATATTTCTGTGCTGTTTGTGCATATATAAACCCTCAAA 325  
|||||

QY 301 ACTATATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATATACCTTTCTAC 360  
|||||  
Db 326 ACTATATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATATACCTTTCTAC 385  
|||||

QY 361 TTGCGATGATTAACAAAGCAAGCTGACGACGACGACGACGACGACGACGACGACGACG 420  
|||||  
Db 386 TTGCGATGATTAACAAAGCAAGCTGACGACGACGACGACGACGACGACGACGACGACG 445  
|||||

QY 421 CAGGCTAGTAGTAAGTTTGGTCTGCTGCTAGGAAAGGGTC 460  
|||||  
Db 446 CAGGCTAGTAGTAAGTTTGGTCTGCTGCTAGGAAAGGGTC 485  
|||||

RESULT 5

AI221902 406 bp mRNA linear EST 30-NOV-1998

LOCUS q99h07.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

DEFINITION IMAGE:1843357 3', mRNA sequence.

ACCESSION AI221902

VERSION AI221902.1 GI:3804105

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 406)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 902 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 394.

Location/Qualifiers

1. 406

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1843357"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHR, and B-cell NCI-CGAP\_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 146 a 62 c 52 g 146 t

ORIGIN

Query Match 74.2%; Score 370.8; DB 9; Length 406;  
Best Local Similarity 98.7%; Pred. No. 6.5e-71;  
Matches 385; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 1 GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTAAGAGA 60  
|||||  
Db 17 GTATATGTAAGAAGCCTCATCTTTTGAATTTTAAATATACAAGATGCTTCTTTAAGAGA 76  
|||||

QY 61 GCAAGATTCAAATGTTTGTCTTTCACAAATTTAAATATAATATATCTCTTAAATTTT 120  
|||||  
Db 77 GCAAGATTCAAATGTTTGTCTTTCACAAATTTAAATATAATATATCTCTTAAATTTT 136  
|||||

QY 121 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 180  
|||||  
Db 137 CTAAGACATGTTTCATATATTTGACCAATCCCTTATTTGGCAAGGATTTTAAAGAGTCT 196  
|||||

QY 181 AACTCAAAACATATGTAAGCTCTGCTGTACCTGTTATATATATATATATATATATATAT 237  
|||||  
Db 197 AACTCAAAACATATGTAAGCTCTGCTGTACCTGTTATATATATATATATATATATATAT 256  
|||||

QY 238 GATCTATATACATAGACATGAATATATTTCTGTGCTGTTTGTGCATATATAAACCCTCA 297  
|||||  
Db 257 GATCTATATACATAGACATGAATATATTTCTGTGCTGTTTGTGCATATATAAACCCTCA 316  
|||||

QY 298 AACACTATATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATATATACCTTTC 357  
|||||  
Db 317 AACACTATATTAATGCAATCCCTATATTTCTAGGTATAGAAAGTTCATATATACCTTTC 376  
|||||

QY 358 TACTTCCATGGCATTAAACAAAGCAAGGCT 387  
|||||

Db 377 TACTTCCATGGCATTAAACAAAGCAAGGCT 406  
|||||

RESULT 6

BM144799 467 bp mRNA linear EST 30-NOV-2001

LOCUS TCAAPID13102 Pediatric acute myelogenous leukemia cell (FAB M1),

DEFINITION Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP1310, mRNA sequence.

ACCESSION BM144799

VERSION BM144799.1 GI:17162235

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 467)

Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman, F.R. Jr., Gunaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F. Pediatric Leukemia cDNA Sequencing Project (2001)

Unpublished (2001)

Contact: Dr. Judith F. Margolin

Texas Children's Cancer Center and Human Genome Sequencing Center at Baylor College of Medicine

1102 Bates, MC3-3320 Houston, TX 77030, USA

Tel: 832-824-4536

Fax: 832-825-4038

Email: clones@txccc.org

FEATURES  
source

Seq primer: M13 primer.  
Location/Qualifiers  
1. .467  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TCAAP1310"  
/clone\_lib="Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project-TCAA"  
/sex="male"  
/tissue\_type="leukopheresis"  
/cell\_type="myeloid cell"  
/dev\_stage="pediatric 6 years"  
/lab\_host="DH10B"  
/note="vector: lambda pSB; Site\_1: BamHI; Site\_2: EcoRI; First strand cDNA was primed with an anchored xhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCGCAGGAG(T)VN 3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand was primed with a BamHI-dc primer [5'AGAGAGTCGATCCGGCGCGCAATAATAAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda pSB vector. Library went through one round of normalization. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)."  
BASE COUNT 159 a | 65 c. 180 t | 1 others  
ORIGIN

Query Match 50.2%; Score 250.8; DB 13; Length 467;  
Best Local Similarity 97.8%; Pred. NO. 8.7e-45;  
Matches 266; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
Qy 1 GTATATGTAGAAGCCGTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 60  
Db 195 GTATATGTAGAAGCCGTCATCTTTTGGATTTTAAATATACAAGATGCTTTCTTTAAGAGA 254  
Qy 61 GCAAGATTCAAATGTTTGTGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 120  
Db 255 GCAAGATTCAAATGTTTGTGTTTCAAAATTTAAATATATATCTCTCTAAATTTT 314  
Qy 121 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGTCT 180  
Db 315 CTAAGACATGTTTCATATATTTGACCATCCCTTATTTTGGCAAGGATTTTAAAGTCT 374  
Qy 181 AACTCAACATATGTAGCTCTGCTGACCTGGTTATATATACC----AAAAAACAATT 236  
Db 375 AACTCAACATATGTAACTCTGCTGACCTGGTTATATATATACCAAAAAAACAATT 434  
Qy 237 TGATCTATATACATACATGATGAATATATTT 268  
Db 435 TGAATCTATATACATACATGATGAATATATTT 466

RESULT 7  
AW968905/c  
LOCUS AW968905 560 bp mRNA linear EST 01-JUN-2000  
DEFINITION EST380981 MAGE resequences, MAGJ Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW968905  
VERSION AW968905.1 GI:8158746  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Hegde,P., Oi,R., Abernathy,K., Dharap,S., Gaspar,D., Gay,C., Holt I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.  
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL  
COMMENT

Unpublished (2000)  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 258

Seq primer: Forward.

FEATURES  
source

Location/Qualifiers  
1. .560  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGJ"  
/note="vector: pbluescriptSKm"  
BASE COUNT 163 a | 122 c | 120 g | 155 t  
ORIGIN

Query Match 48.9%; Score 244.4; DB 10; Length 560;  
Best Local Similarity 92.4%; Pred. NO. 2.1e-43;  
Matches 268; Conservative 0; Mismatches 21; Indels 1; Gaps 1;  
Qy 212 GGTATATATACCAAAAAACATTTGATCTATATACATACATGATATAT-TTCT 270  
Db 558 GGTGACATATAGGAAGCAAAACATTTGATCTAGATCTATCAGAGCGGATATATAT 499  
Qy 271 GTGTGTGTTTGTGCATATATAACCTCAACACTATTTAAATGCAATCCTATATCTTA 330  
Db 498 GTGTGTGTTTGGCAAGTATAACCTCAAGCACTATAAGTAATGCAATCCTATATCTTA 439  
Qy 331 GGTATAGAAGTTGATGATATACCTTTTCTTACTTGGCATGGCAATTAACAAAGAGCTGAG 390  
Db 438 GGTATAGAAGTTGATGATATAACCTTTTGTACTTGGCATGGCAATTAACAAAGAGCTGAG 379  
Qy 391 ACTCAGCAACCACTTGTTTCATTCGCAATGCAAGCTAGTAGTAGTTGGTTGCTGGTAG 450  
Db 378 ACTCAGCAACCACTTGTTTCATTCGCAATGCAAGCTAGTAGTAGTTGGTTGCTGGTAG 319  
Qy 451 GAAAAGGTCCTTATCTCACCCTCTTAAACTAAAGGTTCTTCAGGCT 500  
Db 318 GAAAAGGTCCTTATCTCACCCTCTTAAACTAAATGTTGTTCTTCAGGCT 269

RESULT 8  
AI570720 217 bp mRNA linear EST 14-MAY-1999  
LOCUS tr56a05.x1 NCI\_CGAP\_Paol Homo sapiens cDNA clone IMAGE:2223248 3',  
DEFINITION mRNA sequence.  
ACCESSION AI570720  
VERSION AI570720.1 GI:4534094  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 217)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
COMMENT Unpublished (1997)  
Tumor Gene Index  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1407 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 186  
POLYA-No.

FEATURES  
source

Location/Qualifiers  
1. .217



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:223248"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
```

BASE COUNT	77 a	30 c	29 g	81 t
ORIGIN				
Query Match	40.5%; Score 202.4; DB 9; Length 217;			
Best Local Similarity	99.5%; Pred. No. 3.2e-34;			
Matches 203; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			

QY 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 60  
|||||  
Db 14 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 73  
|||||  
QY 61 GCAAGATTCAAAATCTTTGCTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 120  
|||||  
Db 74 GCAAGATTCAAAATCTTTGCTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 133  
|||||  
QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTGGCAAGGATTTTAAAGAGTCT 180  
|||||  
Db 134 CTAAGACATGTTTCATATATTGACCATCCCTTATTGGCAAGGATTTTAAAGAGTCT 193  
|||||  
QY 181 AACTCAACATATGTAAGCTCTGG 204  
|||||  
Db 194 AACTCAACATATGTAAGCTCTGG 217  
|||||

RESULT 9  
AA904948  
LOCUS  
DEFINITION  
oJ81b10.s1 Soares\_NFL\_T\_GBC.S1 Homo sapiens cDNA clone  
IMAGE:1504699 3', mRNA sequence.  
AA904948  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 318)  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1030 Std Error: 0.00  
Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 152.  
Location/Qualifiers  
1. .318  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1504699"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19w, testis NHT, and B-cell  
NCI-CGAP.GC81) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.

BASE COUNT	101 a	47 c	47 g	123 t
ORIGIN				
Query Match	30.9%; Score 154.6; DB 9; Length 318;			
Best Local Similarity	97.5%; Pred. No. 8.1e-24;			
Matches 157; Conservative	0; Mismatches 4; Indels 0; Gaps 0;			

QY 1 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 60  
|||||  
Db 158 GTATATGTAAGAAGCCTCATCTTTTGATTTTAAATATACAGATGCTTCTTTAAGAGA 217  
|||||  
QY 61 GCAAGATTCAAAATCTTTGCTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 120  
|||||  
Db 218 GCAAGATTCAAAATCTTTGCTTTCAAAATTTAAATAAATTTATCTCTAAATTTT 277  
|||||  
QY 121 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGG 161  
|||||  
Db 278 CTAAGACATGTTTCATATATTGACCATCCCTTATTTTGG 318  
|||||

RESULT 10  
AA736766  
LOCUS  
DEFINITION  
oa23a03.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1305772 3',  
mRNA sequence.  
AA736766  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
EST.  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 317)  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html  
Insert Length: 1055 Std Error: 0.00  
Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 252.  
Location/Qualifiers  
1. .317  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1305772"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-NGTTACCAATCTGAGTGGAGCGCGCTCATTTTTTTTTTTT-3'  
]. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I



and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 100 a 45 c 44 g 128 t  
ORIGIN  
Query Match 29.9%; Score 149.4; DB 9; Length 317;  
Best Local Similarity 99.3%; Pred. NO. 1.le-22;  
Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAAGATGCTTTCTTAAAGAGA 60  
Db 167 GTATATGTAAGAACCCCTCATCTTTTGATTTTAAATATACAAAGATGCTTTCTTAAAGAGA 226  
QY 61 GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATAATTAATATCTCTTAATTTT 120  
Db 227 GCAAGATTCAAATTTGTTTGTGTTTCAAAATTTAAATAATTAATATCTCTTAATTTT 286  
QY 121 CTAAGAGATGTTTCATATATTTGACCATCC 151  
Db 287 CTAAGAGATGTTTCATATATTTGACCATCC 317

RESULT 11  
AA488780/c  
LOCUS AA488780 360 bp mRNA linear EST 15-AUG-1997  
DEFINITION aa54c10.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:824754 5',  
mRNA sequence.  
ACCESSION AA488780  
VERSION AA488780.1 GI:2218382  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 358.  
Location/Qualifiers  
1 .360  
/organism="Homo sapiens"  
/db.xref="taxon:9606"  
/clone="IMAGE:824754"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD+),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3',  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was

FEATURES  
source

BASE COUNT 108 a 77 c 86 g 89 t  
ORIGIN  
Query Match 26.4%; Score 131.8; DB 9; Length 360;  
Best Local Similarity 98.5%; Pred. NO. 7.4e-19;  
Matches 133; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 366 ATGGCATTAAACAAAGCGTGTGAGCTCAGCAACCACTGTGTTCATTGTCATTGCAGGC 425  
Db 360 ATGGCATTAAACAAAGCGTGTGAGCTCAGCAACCACTGTGTTCATTGTCATTGCAGGC 301  
QY 426 TAGTAGTAGTGTGTTGCTCTGTAGGAAAAGGCTCTTATCTATCTACCCCTCTTAAACATA 485  
Db 300 TAGTAGTAGTGTGTTGCTCTGTAGGAAAAGGCTCTTATCTATCTACCCCTCTTAAACATA 241  
QY 486 AGGTTCTTTCAGGCT 500  
Db 240 TGGTTCTTTCAGGCT 226  
RESULT 12  
AK003359/c  
LOCUS AK003359 1796 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus 18 days embryo, whole body cDNA, RIKEN full-length  
enriched library, clone:1110003109:homolog to SPINAL-CORD DERIVED  
GROWTH FACTOR-B, full insert sequence.  
ACCESSION AK003359  
VERSION AK003359.1 GI:12833975  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,  
clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:1110003109.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multipipillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gliss, C., King, B., Kochiwa, H.,  
Kuenli, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,



[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 4, 2003, 19:30:58 ; Search time 146 Seconds  
(without alignments)  
12059.771 Million cell updates/sec

US-10-083-853B-2

Sequence: 1 gttatgtaagaagcctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62  
Xgapext 0.0, Xgapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10083853/runat\_03072003\_093612\_8118/app-query.fasta\_1.30087  
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10083853.scgn.1.1.221.runat.03072003\_093612\_8118 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGUEURY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	409	0.8	370	4	US-09-457-066-37
C 2	409	0.8	370	4	Sequence 37, Appli
C 3	382	0.7	370	4	Sequence 2, Appli
C 4	314	0.6	1079	3	US-09-540-224-2
C 5	311	0.6	132	4	US-09-540-224-4
C 6	306	0.6	375	2	US-09-058-489-22
C 7	306	0.6	375	2	Sequence 22, Appli
C 8	306	0.6	375	2	Sequence 573, App
C 9	280	0.5	397	5	Sequence 121, App
C 10	271.5	0.5	110	4	Sequence 121, App
C 11	272	0.5	1350	4	PCT-US95-17111A-121
C 12	253	0.5	122	4	Sequence 193, App
					Sequence 17, Appli
					Sequence 285, App

C 13	239	0.5	345	4	US-09-040-2202-2
C 14	239	0.5	345	4	US-09-457-066-2
C 15	239	0.5	345	4	US-09-265-686-2
C 16	239	0.5	345	4	US-09-540-224-5
C 17	235	0.4	99	4	US-09-288-143-168
C 18	235	0.4	345	4	US-09-457-066-43
C 19	209.5	0.4	10182	4	US-09-134-001C-3159
C 20	198	0.4	90	4	US-09-227-357-171
C 21	194.5	0.4	1312	2	US-09-227-357-577
C 22	185.5	0.4	1312	2	US-08-592-126-148
C 23	185.5	0.4	1312	2	US-08-687-080-51
C 24	184.5	0.4	1350	4	US-09-245-041-17
C 25	183	0.3	1588	5	PCT-US93-07261-11
C 26	183	0.3	1663	5	PCT-US93-07261-16
C 27	181.5	0.3	76	4	US-09-605-785-575
C 28	178	0.3	3248	1	US-08-353-700-1
C 29	178	0.3	3248	5	PCT-US95-16216-1
C 30	177	0.3	2482	1	US-08-328-254-6
C 31	173	0.3	500	4	US-09-265-630-13
C 32	170.5	0.3	52	4	US-09-227-357-537
C 33	168	0.3	388	4	US-09-265-630-11
C 34	166	0.3	72	4	US-09-227-357-655
C 35	165.5	0.3	3200	2	US-08-477-451-8
C 36	160	0.3	3287	2	US-08-477-451-7
C 37	156.5	0.3	1388	4	US-09-572-191-2
C 38	156.5	0.3	1388	4	US-09-723-262-2
C 39	156.5	0.3	1388	4	US-09-723-219-2
C 40	156.5	0.3	1886	4	US-08-938-105-3
C 41	146.5	0.3	944	4	US-09-134-001C-4352
C 42	146.5	0.3	976	4	US-09-104-324B-4
C 43	144.5	0.3	1104	4	US-08-923-992A-4
C 44	144.5	0.3	1354	3	US-08-685-871-2
C 45	145	0.3	1388	2	US-08-685-576-1

ALIGNMENTS

RESULT 1

US-09-457-066-37

Sequence 37, Application US/09457066

Patent No. 5,326,731

GENERAL INFORMATION:

APPLICANT: Gao, Zeren

APPLICANT: Hart, Charles E.

APPLICANT: Piddington, Christopher S.

APPLICANT: Sheppard, Paul O.

APPLICANT: Shoemaker, Kimberly E.

APPLICANT: Gilbertson, Debra G.

APPLICANT: West, James W.

TITLE OF INVENTION: GROWTH FACTOR, HOMOLOG ZVEGF3

FILE REFERENCE: 98/450,441

CURRENT APPLICATION NUMBER: US/09/457-066

CURRENT FILING DATE: 1999-12-07

NUMBER OF SEQ ID NOS: 50

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 37

LENGTH: 370

TYPE: PRT

ORGANISM: Homo sapiens

US-09-457-066-37

Alignment Scores:

Pred. No.: 6.98e-32

Score: 409-00

Percent Similarity: 100-00%

Best Local Similarity: 100-00%

Query Match: 0

DB: 0

US-10-083-853B-2 (1-29921) x US-09-457-066-37 (1-370)

Qy 17952 GTTGACCTGGATAGCTCAATGATGCCAAGCGTTACAGTTGCTCCAGCAATTAC 17893

|||||



```
QY 10418 CCACCTCCGGGTTCAGTGATTCTCTCCCTCAGCTCCCAAGTAGCTGGACTACAGG 10477
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1009 ProProGlyPheLysArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArg 1028
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10478 CAGGTGCCACCGCTGCTTAATTTTCTGATTTAGTAGAGATGGGTTTCCCGACGTT 10537
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1029 HisLeuProSerCysProThrAsnPheCysIlePheValGluThrGlyPheHisVal 1048
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10538 GCGCGGCTGGTCTTGAACCTCCTGACCTCAGCTGATCTCGCTCAGCCCTCCCAAAAT 10597
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1049 GlyGlnAlaCysLeuGlnLeuLeuThrSerGlyGlyLeuLeuAlaSerAlaSerGlnSer 1068
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10598 GCTGGATTACAGGATGACCGCACTGCGCCCGG 10630
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1069 AlaGlyIleThrGlyValSerHisAlaArg 1079
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 5
US-09-605-785-573
; Sequence 573, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; CURRENT REFERENCE: 210121.427C16
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-573

Alignment Scores:
Pred. No.: 2,62e-22 Length: 132
Score: 311.00 Matches: 66
Percent Similarity: 68.27% Conservative: 5
Best Local Similarity: 63.46% Mismatches: 33
Query Match: 0.59% Indels: 1
DB: 4 Gaps: 0

US-10-083-853b-2 (1-29921) x US-09-605-785-573 (1-132)

QY 10340 TTTTCTTTTTCAGATGGAGTTTCAGTCTCTTGGCCAGGCTAGAGTGAATGCTGTAT 10399
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 28 PhePheLeuArgGlnGlnSerGlyProValAlaGlnAlaGlyValGlnTrpHisasp 47
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10400 CCCGGGTCACTGCAACCTCCACCTCCCGGGTTCAAGTGAATCTCTCGCTCAGCCTCCCA 10459
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeuSerLeuPro 67
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10460 AGTACTGGGACTACAGGACCTGCCACCCCTCGGCTAATTTTGTATTTTAGTAGA 10519
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 68 HisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSerPheSerArg 87
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
QY 10520 GATGGTTT-TCCCCACGTTGGCCAGGCTGTCTTGAACCTCTGACCTCAGGTGATCTGCC 10578
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 88 AspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGlnGlnSerAla 107
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10579 TGCTCAGCTCCCAAAATGCTGGGATTACAGGCATGAGCCACTGCGCCCGGTCTCTTCC 10638
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 CysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisProAlaCysHis 127
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10639 TAACTTCTAAAC 10650
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 128 IleLeuLeuAsn 131
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 6
US-08-454-557C-121
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread-Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-454-557C-121

Alignment Scores:
Pred. No.: 1,92e-21 Length: 375
Score: 306.00 Matches: 76
Percent Similarity: 71.07% Conservative: 10
Best Local Similarity: 62.81% Mismatches: 30
Query Match: 0.58% Indels: 5
DB: 2 Gaps: 2

US-10-083-853b-2 (1-29921) x US-08-454-557C-121 (1-375)

QY 10328 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 10387
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 195 PheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyVal 214
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10388 CAATGGTGTGATCCCGGGTCACTGCAACCTCCACCTCCCGGGTTCGAAGTGATTCCTCTGC 10447
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 215 GlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyLysLeuPheSerCys 234
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 10448 CTCAGCCTCCCAAGTAGCTGGGACTACAGGCACGCTGCCACCGCTGCTAATTTT--- 10505
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```







; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT  
 ; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY

EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 285  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-227-357-285

Alignment Scores:  
Pred. No.: 122  
Score: 1.85e-16  
Length: 51  
Percent Similarity: 253.00  
Conservative: 74.32%  
Best Local Similarity: 68.92%  
Query Match: 0.48%  
Indels: 0  
Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-227-357-285 (1-122)

QY 10409 CTGCAACCTCCACCTCCGGGTCAAGTCTCTGCTGCTCCAGCTCCAGTAGCTGG 10468  
Db 1 MetGlnAlaLeuProGlyPheLysGlnPheSerCysLeuSerLeuProSerArgTtp 20  
QY 10469 GACTACAGGACGTGCCACCCCTGGCTAATTTTGTATTTTAGTAGAGATGGGTTT 10528  
Db 21 AspTyrGlyCysAlaThrGlnHisProAlaAsnPheCysIlePheArgAspArgVal 40  
QY 10529 CCCACGTGGCCAGCGCTGGTCTGAACCTCCAGCTCAGGTGATCTGCTCGCTCAGCC 10589  
Db 41 SerHisValGlyGlnAlaGlyLeuLysLeuLeuThrSerValAspProProAlaTtpAla 60

QY 10589 TCCCAAAATCGTGGGATTACAGGCATGAGCCACTGCGCCGG 10630  
Db 61 SerGlnSerAlaGlyIleThrGlyLysSerHisCysAlaGln 74

## RESULT 13

US-09-040-220D-2  
Sequence 2, Application US/09040220D  
Patent No. 6391311  
GENERAL INFORMATION:  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Kuo, Sophia S.  
TITLE OF INVENTION: NOVEL POLYPEPTIDES HAVING HOMOLOGY TO VASCULAR  
TITLE OF INVENTION: ENDOTHELIAL CELL GROWTH FACTOR AND BONE MORPHOGENETIC  
TITLE OF INVENTION: PROTEIN 1 AND NUCLEIC ACIDS ENCODING SAME, THEIR USES,  
TITLE OF INVENTION: AND PROCESSES FOR THEIR PRODUCTION  
FILE REFERENCE: P1122  
CURRENT APPLICATION NUMBER: US/09/040,220D  
CURRENT FILING DATE: 1998-03-17  
NUMBER OF SEQ ID NOS: 8  
SEQ ID NO 2  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Human  
US-09-040-220D-2

## Alignment Scores:

Pred. No.: 11e-14  
Score: 239.00  
Length: 345  
Percent Similarity: 71.23%  
Conservative: 42  
Best Local Similarity: 57.53%  
Query Match: 0.46%  
Indels: 21  
Gaps: 0

US-10-083-853B-2 (1-29921) x US-09-040-220D-2 (1-345)

QY 17952 GTTGACCTGGATAGCTCAATGATGATGCAAGCTTACAGTTGCCTCCAGGAATTAC 17893  
Db 236 ValAspLeuAsnLeuLeuThrGluValArgLeuTyrSerCysThrProArgAsnPhe 255  
QY 17892 TCGGTCAATATAGACAAGAGCTGAAGTTGGCCAATGTGGTCTCTTCCACGTTGCCCTC 17833  
Db 256 SerValSerIleArgGluLeuLysArgThrAspThrIlePheThrProGlyCysLeu 275  
QY 17832 CTGCTGACGCTGTGAGGAAATCTGGCTGTGGAACCTCACTGGAGGTCTCTGCACA 17773  
Db 276 LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln 295  
QY 17772 TGCATTCAGGGAACCCGTGAAAAAGTATCATGAGGTA 17734  
Db 296 CysValProSerLysValThrLysLysTyrHisGluVal 308

## RESULT 14

US-09-457-066-2  
Sequence 2, Application US/09457066  
Patent No. 6432673  
GENERAL INFORMATION:  
APPLICANT: Gao, Zeren  
APPLICANT: Hart, Charles E.  
APPLICANT: Piddington, Christopher S.  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Shoemaker, Kimberly E.  
APPLICANT: Gilbertson, Debra G.  
APPLICANT: West, James W.  
TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3  
FILE REFERENCE: 98-60  
CURRENT APPLICATION NUMBER: US/09/457,066  
CURRENT FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 345  
TYPE: PRT

; ORGANISM: Homo sapiens  
US-09-457-066-2

## Alignment Scores:

Pred. No.:	1.1e-14	Length:	345
Score:	239.00	Matches:	42
Percent Similarity:	71.23%	Conservative:	10
Best Local Similarity:	57.53%	Mismatches:	21
Query Match:	0.46%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853B-2 (1-29921) x US-09-457-066-2 (1-345)

Qy	17952	GTTCACCTGGATAGCTCAATGATGCCAAGCGTTACAGTTGCACCTCCAGGAATTAC	17893
Db	236	ValAspLeuAsnLeuLeuThrGluGluValArgLeuTyrSerCysThrProArgAsnPhe	255
Qy	17892	TCGGTCAATATAAGAGAGAGCTGAAGTTGGCCAAATGGTGCTCTTCCACAGTTGCCTC	17833
Db	256	SerValSerIleArgGluGluLeuLysArgThrIlePheTrpProGlyCysLeu	275
Qy	17832	CTCGTCAGCGCTGTGGAGGAAATGTGGCTGTGGAACTGTCAACTGGAGGTCCTCGACA	17773
Db	276	LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln	295
Qy	17772	TGCAATTCAGGAAACCGTGAAGAAAGTATCATGAGGTA	17734
Db	296	CysValProSerLysValThrLysLysTyrHisGluVal	308

## RESULT 15

US-09-265-686-2  
; Sequence 2, Application US/09265686  
; Patent No. 6455283  
; GENERAL INFORMATION:  
; APPLICANT: Ferrara, Napoleone  
; TITLE OF INVENTION: POLYPEPTIDES HOMOLOGOUS TO VEGF AND BMP1  
; FILE REFERENCE: P1122P2  
; CURRENT APPLICATION NUMBER: US/09/265,686  
; PRIOR FILING DATE: 1999-03-10  
; PRIOR APPLICATION NUMBER: US 09/040,220  
; PRIOR FILING DATE: 1998-03-17  
; PRIOR APPLICATION NUMBER: US 09/184,216  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 8  
; SEQ ID NO 2  
; LENGTH: 345  
; TYPE: PRT  
; ORGANISM: Human  
US-09-265-686-2

## Alignment Scores:

Pred. No.:	1.1e-14	Length:	345
Score:	239.00	Matches:	42
Percent Similarity:	71.23%	Conservative:	10
Best Local Similarity:	57.53%	Mismatches:	21
Query Match:	0.46%	Indels:	0
DB:	4	Gaps:	0

US-10-083-853B-2 (1-29921) x US-09-265-686-2 (1-345)

Qy	17952	GTTCACCTGGATAGCTCAATGATGCCAAGCGTTACAGTTGCACCTCCAGGAATTAC	17893
Db	236	ValAspLeuAsnLeuLeuThrGluGluValArgLeuTyrSerCysThrProArgAsnPhe	255
Qy	17892	TCGGTCAATATAAGAGAGAGCTGAAGTTGGCCAAATGGTGCTCTTCCACAGTTGCCTC	17833
Db	256	SerValSerIleArgGluGluLeuLysArgThrIlePheTrpProGlyCysLeu	275
Qy	17832	CTCGTCAGCGCTGTGGAGGAAATGTGGCTGTGGAACTGTCAACTGGAGGTCCTCGACA	17773
Db	276	LeuValLysArgCysGlyGlyAsnCysAlaCysCysLeuHisAsnCysAsnGluCysGln	295

Qy	17772	TGCAATTCAGGAAACCGTGAAGAAAGTATCATGAGGTA	17734
Db	296	CysValProSerLysValThrLysLysTyrHisGluVal	308

Search completed: July 4, 2003, 20:41:23  
Job time : 276 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:09:32 ; Search time 1095.73 Seconds  
(without alignments)  
7405.060 Million cell updates/sec

Title: US-10-083-853B-2\_COPY\_29421\_29921  
Perfect score: 501  
Sequence: I ctttgggttttagacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	470.6	93.9	849	9	AU118935	AU118935
C 2	466.4	93.1	658	13	BM509719	BM509719 ig93b09.Y
C 3	465.8	93.0	750	9	AU119706	AU119706 AU119706
C 4	465.8	93.0	1052	11	BC032807	BC032807 Homo sapi
C 5	465.8	93.0	2776	11	BC028293	BC028293 Homo sapi
C 6	464.2	92.7	760	13	BI092874	BI092874 602857901

7	464.2	92.7	1069	13	BM549258	BM549258 AGENCOURT
8	462.6	92.3	673	17	AG102893	AG102893 Pan trogl
C 9	461	92.0	681	10	AW850541	AW850541 IL3-CT021
C 10	459.4	91.7	602	9	AL711636	AL711636 DKF2p686N
C 11	459.4	91.7	695	17	AQ346406	AQ346406 RPCI11-11
C 12	458.4	91.5	1049	14	BQ071465	BQ071465 AGENCOURT
C 13	458.2	91.5	804	12	BG288152	BG288152 602387957
C 14	457.8	91.4	679	17	AQ491435	AQ491435 RPCI-11-2
C 15	457.2	91.3	656	17	AG077683	AG077683 Pan trogl
C 16	454.8	90.8	579	10	AW845700	AW845700 MR0-CT006
C 17	454.6	90.7	805	12	BG680574	BG680574 602629008
C 18	453.8	90.6	1040	12	BE869508	BE869508 601446175
C 19	453	90.4	650	17	AG047226	AG047226 Pan trogl
C 20	453	90.4	840	12	BG623383	BG623383 602648429
C 21	453	90.4	954	12	AQ748961	AQ748961 HS_5574_A
C 22	452.2	90.3	954	12	BG104752	BG104752 602311892
C 23	452	90.2	717	17	AQ052948	AQ052948 RPCI11-49
C 24	451.2	90.1	2703	17	AF101616	AF101616 AF101616
C 25	450.6	89.9	778	13	BI084663	BI084663 602869851
C 26	449	89.6	876	14	BQ932611	BQ932611 AGENCOURT
C 27	448.4	89.5	736	12	BE788398	BE788398 601480244
C 28	447.8	89.4	789	17	AQ740860	AQ740860 HS_2274_A
C 29	446.8	89.2	522	17	AQ354733	AQ354733 CTBI-EI-
C 30	446.8	89.1	659	17	B83611	B83611 RPCI11-16F1
C 31	446.4	89.1	789	13	BI084336	BI084336 602869851
C 32	445.8	89.0	915	12	BG171891	BG171891 602320924
C 33	445.8	89.0	989	12	BG679572	BG679572 602627615
C 34	445.4	88.9	691	17	AG087253	AG087253 Pan trogl
C 35	445.4	88.9	747	12	BG548021	BG548021 602575934
C 36	445.2	88.9	1821	17	AF101873	AF101873 AF101873
C 37	444.8	88.8	863	13	BI257428	BI257428 602967772
C 38	444	88.6	690	9	AL707561	AL707561 DKF2p6861
C 39	443.2	88.5	908	12	BG353707	BG353707 602563823
C 40	442.4	88.3	820	12	BG621601	BG621601 602617011
C 41	442	88.2	517	17	B64761	B64761 CIT-HSP-200
C 42	441.8	88.2	744	17	AQ393258	AQ393258 CTBI-EI-
C 43	441	88.0	999	14	BM905337	BM905337 AGENCOURT
C 44	440.6	87.9	679	14	BM972622	BM972622 UI-CF-BC1
C 45	440.4	87.9	890	17	AQ739555	AQ739555 HS_5381_B

ALIGNMENTS

RESULT 1	AU118935/C	AU118935	849 bp	linear	EST 01-AUG-2002
LOCUS	AU118935	HEMBAL Homo sapiens	cdna clone	HEMBAL004677 5', mRNA	
DEFINITION	AU118935	sequence.			
ACCESSION	AU118935				
VERSION	AU118935.1	GI:10934170			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.				
TITLE	HRI human CDNA project				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human CDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; CDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.				

JOURNAL COMMENT		Location/Qualifiers	
Unpublished (2000)		1..849	
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue		/organism="Homo sapiens"	
Endocrine Pancreas Consortium		/db_xref="taxon:9606"	
Harvard University, Howard Hughes Medical Institute		/clone="HEMBA1004677"	
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge MA 02138		/clone_lib="HEMBA1"	
Tel: 617-495-1812		/tissue_type="whole embryo, mainly head"	
Fax: 617-495-8557		/dev_stage="embryo, 10 weeks"	
Email: dmelton@biohp.harvard.edu		/note="Vector: pME18SFL3"	
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For Information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)		351 a 190 c 137 g 167 t 4 others	
Seq primer: -40RP from Gibco		BASE COUNT	
High quality sequence stop: 481.		ORIGIN	
FEATURES		Query Match 93.9%; Score 470.6; DB 9; Length 849;	
source		Best Local Similarity 96.2%; Pred. No. 7.5e-106;	
1..658		Matches 482; Conservative 0; Mismatches 19; Indels 0; Gaps 0;	
/organism="Homo sapiens"		1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCTATGCTGAATGGTATTGCGCTG 60	
/db_xref="taxon:9606"			
/clone_lib="HR85 islet"		525 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCTATGCTGAATGGTATTGCGCTA 466	
/tissue_type="Purified pancreatic islet"			
/lab_host="DH10B"		QY 61 GGTTCCTCTAGGGAATTTATGTTAGTCTTAACATTTAAAGTCTTTAATCCATCTTG 120	
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site: 1		Db 465 GGTTCCTCTAGGCTTTTATGTTAGTCTTAACATTTAAATCTTTAATCCATCTTG 406	
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.		QY 121 AATTAATTTTGTATAGGTGTAAAGGATCCAGTCTTCAGCTTTCTACATAGGGCTA 180	
Size selected on agarose gel. Average insert size ~1kb. 5		Db 405 AATTGATTTTGTATAGGTGTAAAGGATCCAGTCTTCAGCTTTCTACATAGGGCTA 346	
XhoI site was destroyed after directional cloning.		QY 181 GCCAGTTTTCTCAGCACCAATTATTAATAAGGGAATCCTTCCCATGCTGTTTCT 240	
Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."		Db 345 GCCAGTTTTCCAGCACCAATTATTAATAAGGGAATCCTTCCCATGCTGTTTCT 286	
BASE COUNT 261 a 142 c 120 g 134 t 1 others		QY 241 CAGTTTTGTCAAAGATCAGATAGTTGATATGCGGCATTAATTCATGAGGGCTCTGTC 300	
ORIGIN		Db 285 CAGTTTTGTCAAAGATCAGATAGTTGATATGCGGCATTAATTCATGAGGGCTCTGTC 226	
Query Match 93.1%; Score 466.4; DB 13; Length 658;		QY 301 TGTTCCATTTGGTGAATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360	
Best Local Similarity 95.6%; Pred. No. 7.6e-105;		Db 225 TGTTCCATTTGATCTATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 166	
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0		QY 361 CTTTGTAGTGTAGTTGAAGTCAGGTAGATGATGCTCCAGCTTTGCTTTGGCTTA 420	
QY 1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCTATGCTCTGAATGGTATTGCGCTG 528		Db 165 CTTTGTAGTGTAGTTGAAGTCAGGTAGTGTGATGCTCCAGCTTTGCTTTGGCTTA 106	
Db 587 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCTATGCTCTGAATGGTATTGCGCTA 528		QY 421 GGATTGACTTTGGCGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480	
QY 61 GGTTCCTCTAGGGAATTTATGTTTGGTCTTAACATTTAAGTCTTTAATCCATCTTG 120		Db 105 GGATTGACTTTGGCAATCGCGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 46	
Db 527 GGTTCCTCTAGGCTTTTATGTTTGGTCTTAACGTNTAAGTCTTTAATCCATCTTG 468		QY 481 CCAATTCGTGAAGAAGTAA 501	
QY 121 AATTAATTTTGTATAGGTGTAAAGGATCCAGTCTTCAGCTTTCTACATAGGGCTA 180		Db 45 CCAATTCGTGAAGAAGTCA 25	
Db 467 AATTAATTTTGTATAGGTGTAAAGGATCCAGTCTTCAGCTTTCTACATAGGGCTA 408		RESULT 2	
QY 181 GCCAGTTTTCTCAGCACCAATTATTAATAAGGGAATCCTTCCCATGCTTTGTTTCT 240		BM509719	
Db 407 GCCAGTTTTCCAGCACCAATTATTAATAAGGGAATCCTTCCCATGCTTTGTTTCT 348		LOCUS	
QY 241 CAGTTTTGTCAAAGATCAGATAGTTGATATGCGGCATTAATTCATGAGGGCTCTGTC 300		ig93b09.y1 HR85 islet Homo sapiens cDNA 5' similar to TR:Q14288	
Db 347 CAGTTTTGTCAAAGATCAGATAATGTAGATATGACGCGTTATTTCTGAGGGCTCTGTC 288		DEFINITION	
QY 301 TGTTCCATTTGGTGAATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360		Q14288 HYPOTHETICAL PROTEIN ;, mRNA sequence.	
Db 287 TGTTCCATTTGATCTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 228		ACCESSION	
QY 361 CCITGTAGTGTAGTTTGAAGTCCAGTACGCTCCAGCTTTCTCTTTGGCTTA 168		BM509719	
Db 421 GGATTGACTTTGGCGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480		VERSION	
QY 481 CCAATTCGTGAAGAAGTAA 501		BM509719.1	
Db 107 CCAATTCGTGAAGAAGTCA 87		EST.	
		SOURCE	
		human.	
		ORGANISM	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
		REFERENCE	
		1 (bases 1 to 658)	
		AUTHORS	
		Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas M., Gibbons,M., McCann,R., Cole,R., Tsagaris,Ishvili,R., Williams,T., Jackson,Y. and Bowers,Y.	
		TITLE	
		Endocrine Pancreas Consortium	

Unpublished (2000)		JOURNAL COMMENT	
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue			
Endocrine Pancreas Consortium			
Harvard University, Howard Hughes Medical Institute			
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138			
Tel: 617-495-1812			
Fax: 617-495-8557			
Email: dmelton@biohph.harvard.edu			
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)			
Seq primer: -40RP from Gibco			
High quality sequence stop: 481.		FEATURES	
Location/Qualifiers			
1..658			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone_lib="HR85 islet"			
/tissue_type="Purified pancreatic islet"			
/lab_host="DH10B"			
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. 5' Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."			
BASE COUNT			ORIGIN
261 a 142 c 120 g 134 t 1 others			
Query Match 93.1%; Score 466.4; DB 13; Length 658;			
Best Local Similarity 95.6%; Pred. No. 7.6e-105;			
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;			
QY	1 CTTTGGTGGTTTACACATGAAGTCCTTGCCCATGCTATGCTCAATGGTATTCGCTG 60	QY	
Db	587 CTTTGGTGGTTTACACATGAAGTCCTTGCCCATGCTATGCTCAATGGTATTCGCTA 528		
QY	61 GSTTTTCTCTAGGGATTATGTTTGGTCTTAACATTTAAGTCTTTAATCCATCTTG 120		
Db	527 GSTTTTCTCTAGGGTTTTATGTTTGGTCTTAAGTNTAAGTCTTTAATCCATCTTG 468		
QY	121 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180		
Db	467 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 408		
QY	181 GCCAGTTTCTCAGCACCAATTATTAAATAGGGAATCCTTCCCATGCTGTTTCT 240		
Db	407 GCCAGTTTCCAGCACCAATTATTAAATAGGGAATCCTTCCCATGCTGTTTCT 348		
QY	241 CAGGTTTGTCAAAGATCAGATAGTTGTATATGCGGCATTAATTCATGAGGGCTCTGTTTC 300		
Db	347 CAGGTTTGTCAAAGATCAGATAATGTAGATATGCGGCATTAATTCATGAGGGCTCTGTTTC 288		
QY	301 TGTTCATTTGGTGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360		
Db	287 TGTTCATTTGATCTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 228		
QY	361 CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTCTTTTGGCTTA 420		
Db	227 CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTCTTTTGGCTTA 168		
QY	421 GGATTGACTTTGGCGATGTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480		
Db	167 GGATTGACTTTGGTGTATGCGGCTCTTTTGGTTCATATGAACCTTTGAAAGTAGTTTTT 108		
QY	481 CCAATTCGTGAAGAAGTAA 501		
Db	107 CCAATTCGTGAAGAAGTCA 87		



```

RESULT 3
AUI19706/c
LOCUS
DEFINITION
AUI19706 HEMBA1 Homo sapiens cDNA clone HEMBA1006448 5', mRNA
sequence:
AUI19706
ACCESSION
AUI19706
VERSION
AUI19706.1 GI:10934941
KEYWORDS
EST.
SOURCE
human...
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 750)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1006448"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="vector: pME18SFL3"
BASE COUNT 302 a 166 c 123 g 156 t 3 others
ORIGIN
Query Match 93.0%; Score 465.8; DB 9; Length 750;
Best Local Similarity 95.6%; Pred. No. 1.1e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCATGCGCTATGCTCCTGAATGTCCTG 60
Db 574 CTTTGTGTTTGTGACATGAAGTCCTTGCCAGCGCTATGCTCCTGAATGTCCTA 515
Qy 61 GGTTCCTCTAGGCAATTTATGTTTGTAGTCTAACATTTAAGTCTTTAATCCATCTG 120
Db 514 GGTTCCTCTAGGCTTTTATGTTTGTAGTTTAAAGTTTAAATCTTTAATCCATCTG 455
Qy 121 AATTAATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
Db 454 AATTGATTTTGTATAAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 395
Qy 181 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 240
Db 394 GCCAGTTTCTCAGCACCATTTATTAATAGGGAATCCTTTCCCATTCGTTGTTTCT 335
Qy 241 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 300
Db 334 CAGGTTTGTCAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTC 275
Qy 301 TGTTCCATGTTGGTATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 360
Db 274 TGTTCCATGATCATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACTGTAG 215
Qy 361 CTTTGTAGTTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTGTCTCTTTGGCTTA 420
Db 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCATGCGCTATGCTCCTGAATGTCCTG 60

214 CTTTGTAGTATAGTTTGAAGTCAGGTAGTGTGATGCCCTCCAGCTTGTCTTTGGCTTA 155
Qy 421 GGATTGACTTGGCGATGTGGCTCTTTTTCCTCCACATGAACATTTAAAGTAGTTTTT 480
Db 154 GGATTGACTTGGCGAATGCGGCTCTTTTTCCTCCATATGACATTTAAAGTAGTTTTT 95
Qy 481 CCAATTCTGTGAAGAAAGTAA 501
Db 94 CCAATTCTGTGAAGAAAGTCA 74

RESULT 4
BC032807/c
LOCUS
DEFINITION
BC032807 Homo sapiens, clone IMAGE:4480714, mRNA.
ACCESSION
BC032807
VERSION
BC032807.1 GI:21618542
KEYWORDS
HTC.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1052)
Strausberg,R.
Direct Submission
TITLE
JOURNAL
COMMENT
NH-MGC Project-URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@hri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.S., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgou,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 68 Row: f Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: frame shifted.
FEATURES
Location/Qualifiers
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4480714"
/tissue_type="Bladder, transitional cell papilloma"
/clone_lib="NIH_MGC_93"
/lab_host="DH10B"
/notice="Vector: pCMV-SPORT6"
BASE COUNT 434 a 209 c 182 g 227 t
ORIGIN
Query Match 93.0%; Score 465.8; DB 11; Length 1052;
Best Local Similarity 95.6%; Pred. No. 1.2e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCATGCGCTATGCTCCTGAATGTCCTG 60
```

```

Db      855 CTTTGTGTTTATAGACATGAAGTCCCTGGCCATGCTATGCTCTGAATGTAATGCTTA 796
Qy      61 GGTTCCTCTAGGAGTATTTATGCTTTTAGTCTTAACATTTAAGTCTTTAATCATCTTG 120
Db      795 GGTTCCTCTAGGAGTATTTATGCTTTTAGTCTTAACATTTAAGTCTTTAATCATCTTG 736
Qy      121 AATTAATTTTGTATAGGTGTAAAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGCTTA 180
Db      735 AATTGATTTTGTATAGGTGTAAAGGAAGGATCCAGTTTCAGCTTTTCTACATAGGCTTA 676
Qy      181 GCACGTTTCTCAGCACCATTATTAATAATAGGAATCCCTTCCCATGCTGTTGTTTCT 240
Db      675 GCCAGTTTCTCCAGCACCATTATTAATAATAGGAATCCCTTCCCATGCTGTTGTTTCT 616
Qy      241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTCAGGGCTCTGCTTC 300
Db      615 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTCAGGGCTCTGCTTC 556
Qy      301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTACCATGTTGTTTGGTACTGTAG 360
Db      555 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTACCATGTTGTTTGGTACTGTAG 496
Qy      361 CTTGTAGTGTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGCTTA 420
Db      495 CTTGTAGTGTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGCTTA 436
Qy      421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTGTGTTT 480
Db      435 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTGTGTTT 376
Qy      481 CCAATTCTGTGAAGAAAGTAA 501
Db      375 CCAATTCTGTGAAGAAAGTCA 355

```

RESULT 5  
BC028293/c  
LOCUS  
DEFINITION  
Homo sapiens, clone IMAGE:4704802, mRNA.  
ACCESSION  
BC028293  
VERSION  
BC028293.1 GI:22418059  
KEYWORDS  
HTC.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2776)  
Strausberg,R.  
Direct Submission  
Submitted (10-APR-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 29 Row: 1 Column: 19  
This clone has the following problem: incomplete processing.

```

FEATURES
source
Location/Qualifiers
1..2776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4704802"
/tissue_type="Breast, mammary adenocarcinoma."
/clone_lib="NIH_MGC_87"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
BASE COUNT 1002 a 618 c 587 g 569 t
ORIGIN
Query Match 93.0%; Score 465.8; DB 11; Length 2776;
Best Local Similarity 95.6%; Pred. No. 1.6e-104;
Matches 479; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTATAGACATGAAGTCCCTGGCCATGCTATGCTCTGAATGTAATGCTTA 60
Db 1704 CTTTGTGTTTATAGACATGAAGTCCCTGGCCATGCTATGCTCTGAATGTAATGCTTA 1645
Qy 61 GGTTCCTCTAGGAGTATTTATGCTTTTAGTCTTAACATTTAAGTCTTTAATCATCTTG 120
Db 1644 GGTTCCTCTAGGAGTATTTATGCTTTTAGTCTTAACATTTAAGTCTTTAATCATCTTG 1585
Qy 121 AATTAATTTTGTATAGGTGTAAAGGAAGGATCCAGTTTTCAGCTTTTCTACATAGGCTTA 180
Db 1584 AATTAATTTTGTATAGGTGTAAAGGAAGGATCCAGTTTTCAGCTTTTCTACATAGGCTTA 1525
Qy 181 GCCAGTTTCTCAGCACCATTATTAATAATAGGAATCCCTTCCCATGCTGTTGTTTCT 240
Db 1524 GCCAGTTTCTCAGCACCATTATTAATAATAGGAATCCCTTCCCATGCTGTTGTTTCT 1465
Qy 241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTCAGGGCTCTGCTTC 300
Db 1464 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTCAGGGCTCTGCTTC 1405
Qy 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTACCATGTTGTTTGGTACTGTAG 360
Db 1404 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTACCATGTTGTTTGGTACTGTAG 1345
Qy 361 CTTGTAGTGTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGCTTA 420
Db 1344 CTTGTAGTGTAGTTGAAGTCAAGTATGATGATGCTCCAGCTTTGTTTGGCTTA 1285
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTGTGTTT 480
Db 1284 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTGTGTTT 1225
Qy 481 CCAATTCTGTGAAGAAAGTAA 501
Db 1224 CCAATTCTGTGAAGAAAGTCA 1204

```

RESULT 6  
BI092874/c.  
LOCUS  
DEFINITION  
602857901F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4999176 5',  
mRNA sequence.  
ACCESSION  
BI092874  
VERSION  
BI092874.1 GI:14511204  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 760)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11029 row: a column: 01

High quality sequence stop: 760.

Location/Qualifiers

FEATURES

source

1..760

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4999176"

/clone\_lib="NIH\_MGC\_10"

/cell\_line="MGC36"

/lab\_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5 kb. Library prepared by Life

Technologies."

BASE COUNT 313 a 164 c 133 g 150 t

Query Match 92.7%; Score 464.2; DB 13; Length 760;

Best Local Similarity 95.4%; Pred. No. 2.8e-104;

Matches 478; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATCTCTGAATGGTATTGCCCTG 60

|||||

Db 647 CTTTGGTGTATAGACATGAAGTCTTGGCCGCTATCTCTGAATGGTATTGCCCTA 588

QY 61 GGTTCCTCTAGGATTTATAGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120

|||||

Db 587 GGTTCCTCTAGGATTTATAGTTTATAGTCTTAACGTTTAAGTCTTTAATCCATCTTG 528

QY 121 AATTAATTTTGTATAGTCTTAAGGATGTAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180

|||||

Db 527 AATTAATTTTGTATAGTCTTAAGGATGTAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 468

QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTCT 240

|||||

Db 467 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTCT 408

QY 241 CAGTTTCTCAAGATCAGATGATGTAGATATGCGCATTAATTTCTGAGGCTCTGTTC 300

|||||

Db 407 CAGTTTCTCAAGATCAGATGATGTAGATATGCGCGTATTTCTCAGGCTCTGTTC 348

QY 301 TGTTCATTCGTGATATCTCTGTTTGGTACCATGCTGTTGTTTGGTACTGTAG 360

|||||

Db 347 TGTTCATTCGTGATATCTCTGTTTGGTACCATGCTGTTGTTTGGTACTGTAG 288

QY 361 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 420

|||||

Db 287 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 228

QY 421 GGAATGACTTGGCATGTTGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480

|||||

Db 227 GGAATGACTTGGCATGTTGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 168

QY 481 CCAATTCCTGTAAGAAAGTAA 501

|||||

Db 167 CCAATTCCTGTAAGAAAGTCA 147

RESULT 7

BM549258

LOCUS

DEFINITION BM549258 1069 bp mRNA linear EST 20-FEB-2002

5', mRNA sequence. AGENCOURT\_6558415 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5548510

ACCESSION BM549258

VERSION BM549258.1

KEYWORDS EST. GI:18784517

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 1069)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM12257 row: i column: 23

High quality sequence stop: 695.

Location/Qualifiers

1..1069

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5548510"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2 kb. Library constructed by Life

Technologies."

BASE COUNT 236 a 193 c 224 g 412 t 4 others

ORIGIN

Query Match 92.7%; Score 464.2; DB 13; Length 1069;

Best Local Similarity 95.4%; Pred. No. 3.1e-104;

Matches 478; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATCTCTGAATGGTATTGCCCTG 60

|||||

Db 80 CTTTGGTGTATAGACATGAAGTCTTGGCCATGCCCTATCTCTGAATGGTATTGCCCTA 139

QY 61 GGTTCCTCTAGGATTTATAGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTTG 120

|||||

Db 140 GGTTCCTCTAGGATTTATAGTTTATAGTCTTAACATTTAATCCATCTTG 199

QY 121 AATTAATTTTGTATAGTCTTAAGGATGTAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180

|||||

Db 200 AATTAATTTTGTATAGTCTTAAGGATGTAAGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 259

QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTCT 240

|||||

Db 260 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTCT 319

QY 241 CAGTTTCTCAAGATCAGATGATGTAGATATGCGCATTAATTTCTGAGGCTCTGTTC 300

|||||

Db 320 CAGTTTCTCAAGATCAGATGATGTAGATATGCGGTATTAATTTCTGAGGCTCTGTTC 379

QY 301 TGTTCATTCGTGATATCTCTGTTTGGTACCATGCTGTTGTTTGGTACTGTAG 360

|||||

Db 380 TGTTCATTCGTGATATCTCTGTTTGGTACCATGCTGTTGTTTGGTACTGTAG 439

QY 361 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 420

|||||

Db 440 CTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTGTTCTTTGGCTTA 499

QY 421 GGAATGACTTGGCATGTTGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480

|||||

Db 500 GGAATGACTTGGCATGTTGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 559

QY 481 CCAATTCCTGTAAGAAAGTAA 501

|||||

Db 560 CCAATTCCTGTAAGAAAGTCA 580

## RESULT 8

AG102893 673 bp DNA linear GSS 03-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-106F11.F, genomic survey sequence.  
DEFINITION AG102893  
ACCESSION AG102893  
VERSION AG102893.1 GI:16723410  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
ORGANISM Pan troglodytes

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

## REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

## AUTHORS

Totoki, Y., Watanabe, H. and Sakaki, Y.

## TITLE

BAC end sequences of Library PTB

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 673)

## AUTHORS

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

## TITLE

Direct Submission

## JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpbes@sc.riken.go.jp, URL: http://ngp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

## PRIMERS

Sequencing: -21M13

## LIBRARY

Vector : pKS145

R-Site 1 : SacI

R-Site 2 : SacI

Location/Qualifiers

1. 673

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-106F11.F"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

137 a 119 c 152 g 262 t 3 others

## BASE COUNT

## ORIGIN

Query Match 92.3%; Score 462.6; DB 17; Length 673;

Best Local Similarity 95.2%; Pred. No. 6.7e-104;

Matches 477; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

1 CTTTGGTGTTTAGACATGAAGTCCTTGCCTATGCTATGCTCCTGAATGATTCGCTA 195

136 CTTTGGTGTTTAGACATGAAGTCCTTGCCTATGCTATGCTCCTGAATGATTCGCTA 195

61 GGTTCCTCTAGGATTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 120

196 GGTTCCTCTAGGATTTTATGGTTTATAGTCTAACATTTAAGTCTTTAATCCATCTTG 255

121 AATTAATTTTGTATAGCTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTA 180

256 AATTAATTTTGTATAGCTGTAAGGAGGATCCAGTTTCAGCTTCTACATAGGCTA 315

181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCCTTCCCATTCCTGTTTCT 240

316 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCCTTCCCATTCCTGTTTCT 375

241 CAGGTTTGTAAAGATCAGATAGTTGATAGTGGGCAATTTATCTAGAGGCTCTGTTC 300

376 CAGGTTTGTAAAGATCAGATAGTTGATAGTGGGCAATTTATCTAGAGGCTCTGTTC 435

301 TGTTCATGTTGTATCTCTGTTTGGTACCAGTACCACCATCTGTTTGGTACTGTAG 360

436 TGTTCATGTTGTATCTCTGTTTGGTACCAGTACCACCATCTGTTTGGTACTGTAG 495

## QY

361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTGTCTTTTGGCTTA 420

## Db

496 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCCGCTTGTCTTTTGGCTTA 555

## QY

421 GGATTCACATTCGGCATGTGGGCTCTTTTGGTTCACATGAACACTTAAAGTAGTATTTT 480

## Db

556 GGATTCACATTCGCAATGCAAGTCTCTTTTGTTCATATGAACACTTAAAGTAGTATTTT 615

## QY

481 CCAATTCCTCTGAAGAAAGTAA 501

## Db

616 CCAATTCCTCTGAAGAAAGTAA 636

## RESULT 9

## LOCUS

AW850541/C

## DEFINITION

IL3-CT0219-280100-062-F07 CT0219 Homo sapiens cDNA, mRNA sequence.

## ACCESSION

AW850541

## VERSION

AW850541.1

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 681)

## AUTHORS

Dias Neto, E., Garcia

Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

## JOURNAL

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## MEDLINE

20202663

## COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6&2=IL3-CT0219-280

100-062-F07&t3=2000-01-28&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 5

High quality sequence stop: 535.

## FEATURES

## source

1. 681

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0219"

/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the puc 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 270 a 148 c 122 g 141 t

## ORIGIN

Query Match 92.0%; Score 461; DB 10; Length 681;

Best Local Similarity 95.0%; Pred. No. 1.7e-103;

Matches 476; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

1 CTTTGGTGTTTAGACATGAAGTCCTTGCCTATGCTCCTGAATGATTCGCTG 60

```

Db 611 CTTTGTGTTTGTAGACATGAAGTCCCTGTCGCCATGCCATATGCTCTGAATGGTAAAGGCTA 552
Qy 61 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
Db 551 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 492
Qy 121 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 180
Db 491 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 432
Qy 181 GCAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 240
Db 431 GCAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 372
Qy 241 CAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 300
Db 371 CAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 312
Qy 301 TGTTCATTTGTTATATCTCTGTTTGGTACCATGATCCATGATCCATGATCCATGATCCATGATCC 360
Db 311 TGTTCATTTGTTATATCTCTGTTTGGTACCATGATCCATGATCCATGATCCATGATCCATGATCC 252
Qy 361 CTTGTAGTCTAGTTTGAAGTCAAGGTAGCGTATGATGCGGCGTAAATTTCTGAGGCTCTGTTTCT 420
Db 251 CTTGTAGTCTAGTTTGAAGTCAAGGTAGCGTATGATGCGGCGTAAATTTCTGAGGCTCTGTTTCT 192
Qy 421 GGATTCAGTTCAGGATGTCGGCTCTTTTGGTTCACATGAACTTTAAAGTAGTCTTTTCTTCTTCT 480
Db 191 GGATTCAGTTCAGGATGTCGGCTCTTTTGGTTCACATGAACTTTAAAGTAGTCTTTTCTTCTTCT 132
Qy 481 CCAATTCGTGAAGAAAGTAA 501
Db 131 CCAATTCGTGAAGAAAGTCA 111

```

```

RESULT 10
AL711636/c 602 bp mRNA linear EST 22-MAR-2002
LOCUS DKF2p686N0383_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKF2p686N0383 5', mRNA sequence.
ACCESSION AL711636
VERSION AL711636.1 GI:19694991
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Koehler,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.
AUTHORS EST (Koehler,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.)
TITLE Unpublished (1999)
JOURNAL Contact: Koehler K
COMMENT MIPS
Am Klopferpitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BWFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No sl sequence available.
This clone (DKF2p686N0383) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source Location/Qualifiers
1..602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p686N0383"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"

```

```

/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 238 a 132 c 107 g 124 t 1 others
ORIGIN

Query Match 91.7%; Score 459.4; DB 9; Length 602;
Best Local Similarity 94.8%; Pred. No. 4e-103;
Matches 475; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1 CTTTGTGTTTGTAGACATGAAGTCCCTGTCGCCATGCCATATGCTCTGAATGGTAAAGGCTG 60
Db 591 CTTTGTGTTTGTAGACATGAAGTCCCTGTCGCCATGCCATATGCTCTGAATGGTAAAGGCTA 532
Qy 61 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
Db 531 GGTTCCTCTAGGATTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 472
Qy 121 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 180
Db 471 AATTAATTTTGTATAGGTGTAAAGGATCCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 412
Qy 181 GCAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 240
Db 411 GCAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 352
Qy 241 CAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 300
Db 351 CAGTTCCTCAGCACCATTTTAAATAGGAATCCCTTCCCATTCCTTCCCATTCCTTCCCATTCCT 292
Qy 301 TGTTCATTTGTTATATCTCTGTTTGGTACCATGATCCATGATCCATGATCCATGATCCATGATCC 360
Db 291 TGTTCATTTGTTATATCTCTGTTTGGTACCATGATCCATGATCCATGATCCATGATCCATGATCC 232
Qy 361 CTTGTAGTCTAGTTTGAAGTCAAGGTAGCGTATGATGCGGCGTAAATTTCTGAGGCTCTGTTTCT 420
Db 231 CTTGTAGTCTAGTTTGAAGTCAAGGTAGCGTATGATGCGGCGTAAATTTCTGAGGCTCTGTTTCT 172
Qy 421 GGATTCAGTTCAGGATGTCGGCTCTTTTGGTTCACATGAACTTTAAAGTAGTCTTTTCTTCTTCT 480
Db 171 GGATTCAGTTCAGGATGTCGGCTCTTTTGGTTCACATGAACTTTAAAGTAGTCTTTTCTTCTTCT 112
Qy 481 CCAATTCGTGAAGAAAGTAA 501
Db 111 CCAATTCGTGAAGAAAGTCA 91

RESULT 11
AQ346406/c
LOCUS AQ346406
DEFINITION RPC111-116H20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-116H20,
DNA sequence.
ACCESSION AQ346406
KEYWORDS GI:4171302
SOURCE GSS.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 695)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPC111-116H20.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@igr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```

library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1. .695  
/organism="Homo sapiens"  
/db\_xref="GDB:7544347"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-116H20"  
/clone\_lib="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Site\_3: EcoRI; Human Male BAC Library"  
BASE COUNT 285 a 153 c 118 g 139 t  
ORIGIN

Query Match 91.7%; Score 459.4; DB 17; Length 695;  
Best Local Similarity 94.8%; Pred. No. 4.2e-103;  
Matches 475; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CTTTGGTGTAGACATGAGTCTCTGCCATGCTATGCTCCTGAATGGTATTCCTG 60  
DB 654 CTTTGGTGTGTAGACATGAGTCTCTGCCATGCTATGCTCCTGAATGGTATTCCT 595  
QY 61 GGTTCCTCTAGGATTTATGTTTAGGCTACATTTAGTCTTAACTCACTCTG 120  
DB 594 GGTTCCTCTAGGATTTATGTTTAGGCTACATTTAGTCTTAACTCACTCTG 535  
QY 121 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180  
DB 534 AATTGATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 475  
QY 181 GCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCATTCCTGTTTCT 240  
DB 474 GCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCATTCCTGTTTCT 415  
QY 241 CAGTTTGTCAAGATCAGATAGTTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 300  
DB 414 CAGTTTGTCAAGATCAGATAGTTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 355  
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACCTGTC 360  
DB 354 TGTTCATGATCTATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACCTGTC 295  
QY 361 CTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTGTGTTTCTTTGGCTTA 420  
DB 294 CTTGTAGTGTAGTTGAAGTCAGGTAGCTGCTGATGCTCCAGCTTTGTTTGGCTTA 235  
QY 421 GGATTCACCTGGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTCTTTT 480  
DB 234 GGATTCACCTGGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTCTTTT 175  
QY 481 CCAATTTCTGGAAGAAAGTAA 501  
DB 174 CCAATTTCTGGAAGAAAGTCA 154

RESULT 12  
BO071465/c  
LOCUS  
DEFINITION AGENCOURT\_6867166 NIH\_MGC\_47 Homo sapiens cdna clone IMAGE:5927882  
5', mRNA sequence.  
ACCESSION BO071465  
VERSION BO071465.1 GI:19900511  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 1049)  
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ruben Laboratory  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLCM2103 row: 1 column: 03  
High quality sequence stop: 602.

FEATURES  
source  
Location/Qualifiers  
1. .1049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5927882"  
/clone\_lib="NIH\_MGC\_47"  
/tissue\_type="neuroblastoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."  
BASE COUNT 410 a 237 c 194 g 207 t  
ORIGIN

Query Match 91.5%; Score 458.4; DB 14; Length 1049;  
Best Local Similarity 94.6%; Pred. No. 8.2e-103;  
Matches 474; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTAGACATGAGTCTCTGCCATGCTATGCTCCTGAATGGTATTCCTG 60  
DB 721 CTTTGGTGTGTAGACATGAGTCTCTGCCATGCTATGCTCCTGAATGGTATTCCTG 662  
QY 61 GGTTCCTCTAGGATTTATGTTTAGGCTACATTTAGTCTTAACTCACTCTG 120  
DB 661 GGTTCCTCTAGGATTTATGTTTAGGCTACATTTAGTCTTAACTCACTCTG 602  
QY 121 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180  
DB 601 AATTAATTTTGTATAGGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 542  
QY 181 GCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCATTCCTGTTTCT 240  
DB 541 GCAGTTTCTCAGCACCATTATTAATAAGGAAATCCCTTCCCATTCCTGTTTCT 482  
QY 241 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 300  
DB 481 CAGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 422  
QY 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACCTGTC 360  
DB 421 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTACCTGTC 362  
QY 361 CTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTGTGTTTCTTTGGCTTA 420  
DB 361 CTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCCTCAGCTTTGTTTCTTTGGCTTA 302  
QY 421 GGATTCACCTGGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTCTTTT 480  
DB 301 GGATTCACCTGGGATGCGGCTCTTTTGGTTCACATGAACTTAAAGTAGTCTTTT 242  
QY 481 CCAATTTCTGGAAGAAAGTAA 501

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

Db      241 CCAATTCTGTGAAGAAAGTCA 221
|||||
RESULT 13
LOCUS   BG288152/c
DEFINITION 602387957F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516526 5',
mRNA sequence.
ACCESSION BG288152
VERSION   BG288152.1 GI:13042702
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@email.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution by: Incyte Genomics, Inc.
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10408 row: b column: 15
High quality sequence stop: 799.
FEATURES             source
Location/Qualifiers
     1..804
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone_image="4516526"
     /clone_lib="NIH_MGC_93"
     /tissue_type="transitional cell papilloma, cell line"
     /lab_host="DH10B (phage-resistant)"
     /note="Organ: Bladder; Vector: pCMV-SPORT6; Site_1: NotI;
     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
     Average insert size 1.7 kb. Library enriched for
     full-length clones and constructed by Life Technologies.
     Note: this is a NIH_MGC Library."
BASE COUNT 325 a 180 c 135 g 164 t
ORIGIN
Query Match 91.58; Score 458.2; DB 12; Length 804;
Best Local Similarity 96.28; Pred. No. 8.6e-103;
Matches 480; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
Qy 1 CTTTGTGTTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGT-ATTCCCT 59
|||||
Db 615 CTTTGTGTTAGACATCAAGTCCTTGGCCATGCTATGCTCTGAATGGTAAATGCCT 556
|||||
Qy 60 GGGTTTCTCTAGGGATTTTATGGTTAGGTTCAACATTTAAGTCTTTAATCCATCTT 119
|||||
Db 555 AGGTTTCTCTAGGGATTTTATGGTTAGGTTCAACATTTAAGTCTTTAATCCATCTT 496
|||||
Qy 120 GAATTAATTTTGTATTAAGTGTAAAGGAAGGATCCAGTTTCAGCTTTTCATATAGGGCT 179
|||||
Db 495 GAATTAATTTTGTATTAAGTGTAAAGGAAGGATCCAGTTTCAGCTTTTCATATAGGGCT 436
|||||
Qy 180 AGCCAGTTTCTCAGCACCATTATTAAATAGGGAATCCCTTCCCATTCCTGTTTTC 239
|||||
Db 435 AGCCAGTTTCCAGCACCATTATTAAATAGGGAATCCCTTCCCATTCCTGTTTTC 376
|||||
Qy 240 TCAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTT 299
|||||
Db 375 TCAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGGCCTTAATTTCTGAGGGCTCTGTT 316
|||||
Qy 300 CTGTTCCATTTGGTTGATATCTCTGTTTGGTACAGTACATGTTGTTTGGTTACTGTA 359
|||||

```

```

315 CTGTTCCAGTGATCTATATCTCTCTTTTGGTACCAGTACCATGCTGTTTGGTACTGTA 256
Qy 360 GCCTTGTAGTCTAGCTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTCTTTGGCTT 419
|||||
Db 255 GCCTTGTAGTATAGCTTTGAAGTCAGGTAGTGATGCTCCAGCTTTGTTCTTTGGCTT 196
|||||
Qy 420 AGGATTGACCTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTATTT 479
|||||
Db 195 AGGATTGACCTGGCGATATGGCTCTTTTGGTTCCATATGAACCTTTAAAGTAGTATTT 136
|||||
Qy 480 TCCAATTCTGTGAAGAAAG 498
|||||
Db 135 TCCAATTCTGTGAAGAAAG 117
|||||
RESULT 14
LOCUS   AQ491435/c
DEFINITION RPCI-11-244B1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-244B1,
DNA sequence.
ACCESSION AQ491435
VERSION   AQ491435.1 GI:4677309
KEYWORDS GSS.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 679)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-244B1.T3
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             source
Location/Qualifiers
     1..679
     /organism="Homo sapiens"
     /db_xref="GDB:7593336"
     /db_xref="taxon:9606"
     /clones="RPCI-11-244B1"
     /clone_lib="RPCI-11"
     /sex="Male"
     /cell_type="Lymphocytes"
     /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
     RPCI11 Human Male BAC Library"
BASE COUNT 273 a 146 c 118 g 142 t
ORIGIN
Query Match 91.48; Score 457.8; DB 17; Length 679;
Best Local Similarity 94.68; Pred. No. 1e-102;
Matches 474; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 1 CTTTGTGTTTGTAGACATGAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTCGCTG 60
|||||
Db 654 CTTTGTGTTTGTAGATATGAAGTCCTTGGCCATGCTATGCTCTGAATGGTATTCGCTA 595
|||||
Qy 61 GGTTCCTCTAGGGATTTTATGCTTTTAGTCTTAACATTTAAGTCTTTAATCCATCTG 120
|||||

```



```

Db 594 GGTTCCTCTAGGGTTTTATGGTTTAGGCTAACATGTAAGTCCTTTAATCCACCTG 535
Qy 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTTCAGCTTTTCTACATAGGCTA 180
Db 534 AATTAATTTTGTATAGGTGTAAGGAAGGATACAGTTTAACTTTCTACATAGGCTA 475
Qy 181 GCAGTTTCTCAGCACCATTATTAATAAGGAAGAACTTCCCATGCTGCTTTCTTCT 240
Db 474 GCCAGTTTCCAGCACCATTATTAATAAGGAAGAACTTCCCATGCTGCTTTCTTCT 415
Qy 241 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTT 300
Db 414 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTT 355
Qy 301 TGTTCATGTTGTATATCTCTCTTTTGGTACAGTACCAGTGTGTTTGGTTTACGTAG 360
Db 354 TGTTCATGTTGTATATCTCTCTTTTGGTACAGTACCAGTGTGTTTGGTTTACGTAG 295
Qy 361 CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTTCTTTTGGCTTA 420
Db 294 CCTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTTCTTTTGGCTTA 235
Qy 421 GGATTGACTTGGGATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTCT 480
Db 234 GGATTGACTTGGCAATGCAGGCTCTTTTGGTTCACATGAACCTTAAAGTAGTTTCT 175
Qy 481 CCAATCTGTGAAGAAAGTAA 501
Db 174 CCAATCTGTGTAGAAAGTCA 154

RESULT 15
AG077683
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM
Pan troglodytes male lymphoblast DNA, clone: PTB-072E06.F, genomic survey sequence.
GSS.
AG077683.1 GI:16629485

REFERENCE.
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 656)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:champs@sc.riken.go.jp, URL:http://hgw.psc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. .656
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-072E06.F"
/sex="male"
/cell_type="lymphoblast"

FEATURES
source

```

```

BASE COUNT      131 a      118 c      153 g      254 t
ORIGIN
Query Match      91.3%; Score 457.2; DB 17; Length 656;
Best Local Similarity 96.2%; Pred. No. 1.4e-102;
Matches 479; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

Qy 1 CTTTTCGTGTTTGTAGACATGAAGTCCTTGCCTATGCTCCTCAATGATGCTGCTGCTGCTG 60
Db 131 CTTTTCGTGTTTGTAGACATGAAGTCCTTGCCTATGCTCCTCAATGATGCTGCTGCTGCTG 190
Qy 61 GGTTCCTCTCAGGATTTTATGGTTTATAGGTTCTAATTAAGTCTTTAATCCATCTTG 120
Db 191 GGTTCCTCTCAGGATTTTATGGTTTATAGGTTCTAATTAAGTCTTTAATCCATCTTG 250
Qy 121 AATTAATTTTGTATAGGTTGTAAGGAAGGATCCAGTTTTCAGCTTTCTACATAGGCTA 180
Db 251 AATTAATTTTGTATAGGTTGTAAGGAAGGATCCAGTTTTCAGCTTTCTACATAGGCTA 309
Qy 181 GCCAGTTTCTCAGCACCATTATTAATAAGGAAGAACTTCCCATGCTTCTTTTCT 240
Db 310 GCCAGTTTCCAGCACCATTATTAATAAGGAAGAACTTCCCATGCTTCTTTTCT 369
Qy 241 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTT 300
Db 370 CAGGTTTGTCAAAGATCAGATAGTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTT 429
Qy 301 TGTTCATGTTGTATATCTCTCTTTTGGTACAGTACCAGTGTGTTTGGTTTACGTAG 360
Db 430 TGTTCATGTTGTATATCTCTCTTTTGGTACAGTACCAGTGTGTTTGGTTTACGTAG 489
Qy 361 CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTTGTTTCTTTTGGCTTA 420
Db 490 CCTTGTAGTGTAGTTTGAAGTCAGGTAGTGTGATGCTCCAGCTTTGTTTCTTTTGGCTTA 549
Qy 421 GGATTGACTTGGCGATGTGGGCTCTTTTGGTTCCACATGAACCTTAAAGTAGTTTCT 480
Db 550 GGATTGACTTGGCGATGTGGGCTCTTTTGGTTCCATATGAACCTTAAAGTAGTTTCT 609
Qy 481 CCAATTTCTGTGAAGAAAG 498
Db 610 CCAATTTCTGTGAAGAAAG 627

Search completed: July 4, 2003, 17:35:43
Job time : 1099.73 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 16:40:54 ; Search time 112.075 Seconds  
(without alignments)  
6978.003 Million cell updates/sec

Title: US-10-083-853B-2\_COPY\_29421\_29921

Perfect score: 501  
Sequence: 1 ctttgggttttagacatg.....caattctgtgaagaagataa 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1 published Applications\_NA:  
1: /cgn2\_5/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_5/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
3: /cgn2\_5/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_5/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_5/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_5/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
7: /cgn2\_5/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_5/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_5/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_5/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_5/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_5/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_5/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_5/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	501	100.0	29921	9	US-10-083-853-1
2	478.6	95.5	1	9	US-10-050-882-39
3	477.4	95.3	4709	9	US-10-050-882-40
4	472.8	94.4	326014	10	US-09-731-231A-3
5	472.2	94.3	169327	10	US-09-777-921A-3
6	472.2	94.3	167343	10	US-09-962-436-281
7	472.2	94.3	167343	10	US-09-964-824A-273
8	469	93.6	180557	12	US-10-003-806-6
9	469	93.6	180557	12	US-10-003-806-9
10	467.4	93.3	32042	9	US-09-764-891-8497
11	467.4	93.3	32042	9	US-10-118-984-44
12	467.4	93.3	32042	9	US-10-295-981-63
13	467.4	93.3	32042	10	US-09-728-721-63
14	467.4	93.3	32041	9	US-10-006-883A-15
15	467.4	93.3	1691139	9	US-10-067-514-1
16	466	93.0	11057	9	US-09-764-868-1308
17	465.8	93.0	6175	9	US-10-079-854-294
18	465.8	93.0	6175	10	US-09-764-878-294
19	465.8	93.0	7626	9	US-10-001-835-82

ALIGNMENTS

RESULT 1  
US-10-083-853-1  
; Sequence 1, Application US/10083853  
; Patent No. US0020164709A1

; GENERAL INFORMATION:  
; APPLICANT: Affymetrix, Inc  
; APPLICANT: Shigeta, Ron T

; APPLICANT: Siyani-Rose, Michael A  
; TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein

; FILE REFERENCE: 3385.1  
; CURRENT APPLICATION NUMBER: US/10/083,853

; PRIOR FILING DATE: 2002-02-26  
; PRIOR APPLICATION NUMBER: USSN 60/272,663

; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent version 3.1  
; SEQ ID NO 1

; LENGTH: 29921  
; TYPE: DNA

; ORGANISM: Homo Sapiens  
US-10-083-853-1

Query Match 100.0%; Score 501; DB 9; Length 29921;  
Best Local Similarity 100.0%; Pred. No. 1.2e-128;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTTTTGGTCTTTAGACATGAAGTCTTCCCATCGCTATGCTCGAATGTTATTCGCTG	60
DB	29421	CTTTTGGTCTTTAGACATGAAGTCTTCCCATCGCTATGCTCGAATGTTATTCGCTG	29480
QY	61	GGTTTCTTCTAGGGATTTTATGGTCTTAACATTAAGTCTTTAATCCATCTTG	120
DB	29481	GGTTTCTTCTAGGGATTTTATGGTCTTAACATTAAGTCTTTAATCCATCTTG	29540
QY	121	AATTAATTTTGTATAAGGTGTAAGGAAGGATFCCAGTTCAGTTCATACATAGGGCTA	180
DB	29541	AATTAATTTTGTATAAGGTGTAAGGAAGGATFCCAGTTCAGTTCATACATAGGGCTA	29600
QY	181	GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTCCCATTCCTGTTTTCT	240
DB	29601	GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTCCCATTCCTGTTTTCT	29660

QY 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGGCAATATTTCTGAGGGCTCTGTTTC 300  
|||||  
Db 29661 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGGCAATATTTCTGAGGGCTCTGTTTC 29720  
|||||  
QY 301 TGTTCATGCTTGCATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360  
|||||  
Db 29721 TGTTCATGCTTGCATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 29780  
|||||  
QY 361 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTCTTTTGGCTTA 420  
|||||  
Db 29781 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTCTTTTGGCTTA 29840  
|||||  
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 480  
|||||  
Db 29841 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 29900  
|||||  
QY 481 CCAATTCTGTGAAGAAAGTAA 501  
|||||  
Db 29901 CCAATTCTGTGAAGAAAGTAA 29921  
|||||

## RESULT 2

US-10-050-882-39

; Sequence 39, Application US/10050882

; Publication No. US20030104400A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 27 Human secreted proteins

; FILE REFERENCE: P2038P1

; CURRENT APPLICATION NUMBER: US/10/050,882

; CURRENT FILING DATE: 2002-01-18

; PRIOR APPLICATION NUMBER: 09/661,453

; PRIOR FILING DATE: 2000-09-13

; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/125,055

; PRIOR FILING DATE: 1999-03-18

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 39

; LENGTH: 5065

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (2531)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (5063)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-050-882-39

Query Match 95.5%; Score 478.6; DB 9; Length 5065;

Best Local Similarity 95.4%; Pred. No. 8.7e-123;

Matches 478; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTATGCTCCTGAATGGTATTGCTG 60  
|||||Db 1090 CTTTGTGTTTGTAGWCATGAAGTCCTTGCCCATGCTATGCTCCTGAATGGTATTGCTG 1149  
|||||QY 61 GGTTCCTCTAGGATTTTATGGTTTGTAGTCTAACATTTAGTCTTTTAAATCCATGTTG 120  
|||||Db 1150 GGTTCCTCTAGGATTTTATGGTTTGTAGTCTAACATTTAGTCTTTTAAATCCATGTTG 1209  
|||||QY 121 AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTTA 180  
|||||Db 1210 AATTAATTTTGTATAGGTGTAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTTA 1269  
|||||QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATGCTTCTGTTTCT 240  
|||||Db 1270 GCCAGTTTCTCAGCACCATTATTAATAGGGAATCCTTTCCCATGCTTCTGTTTCT 1329  
|||||QY 241 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATCGGGCAATATTTCTGAGGGCTCTGTTTC 300  
|||||

Db 1330 CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGYYGYRTATTTCTGAGGGCTCTGTTTC 1389  
|||||  
QY 301 TGTTCATGCTTGCATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360  
|||||  
Db 1390 TGTTCATGCTTGCATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 1449  
|||||  
QY 361 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTCTTTTGGCTTA 420  
|||||  
Db 1450 CTTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTTCTTTTGGCTTA 1509  
|||||  
QY 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 480  
|||||  
Db 1510 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTT 1569  
|||||  
QY 481 CCAATTCTGTGAAGAAAGTAA 501  
|||||  
Db 1570 CCAATTCTGTGAAGAAAGTAA 1590  
|||||

## RESULT 3

US-10-050-882-40/c

; Sequence 40, Application US/10050882

; Publication No. US20030104400A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 27 Human secreted proteins

; FILE REFERENCE: P2038P1

; CURRENT APPLICATION NUMBER: US/10/050,882

; CURRENT FILING DATE: 2002-01-18

; PRIOR APPLICATION NUMBER: 09/661,453

; PRIOR FILING DATE: 2000-09-13

; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/125,055

; PRIOR FILING DATE: 1999-03-18

; NUMBER OF SEQ ID NOS: 156

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 40

; LENGTH: 4709

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (14)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (18)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (30)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (34)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (177)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (2213)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (3846)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-050-882-40

Query Match 95.3%; Score 477.4; DB 9; Length 4709;

Best Local Similarity 95.4%; Pred. No. 1.8e-122;

Matches 478; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

QY 1 CTTTGTGTTTGTAGACATGAAGTCCTTGCCCATGCTATGCTCCTGAATGGTATTGCTG 60  
|||||Db 3654 CTTTGTGTTTGTAGNCATGAAGTCCTTGCCCATGCTATGCTCCTGAATGGTATTGCTG 3595  
|||||

QY 61 GGTTTCTCTAGGGATTTTATGGTTTATAGTCTAAACATTTAAGTCTTTAATCCACTTG 120  
Db 3594 GGTTTCTCTAGGGATTTTATGGTTTATAGTCTAAACATTTAAGTCTTTAATCCACTTG 3535  
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 180  
Db 3534 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTA 3475  
QY 181 GCCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCT 240  
Db 3474 GCCAGTTTCCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCT 3415  
QY 241 CAGGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGGGCTCTGTT 300  
Db 3414 CAGGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGGGCTCTGTT 3355  
QY 301 TGTTCATCTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 360  
Db 3354 TGTTCATCTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTACTGTAG 3295  
QY 361 CCTGTAGTGTAGTCTGAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 420  
Db 3294 CCTGTAGTATAGTCTGAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTA 3235  
QY 421 GGATTCACCTGGCGATGTGGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTTT 480  
Db 3234 GGATTCACCTGGCGATGTGGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTTT 3175  
QY 481 CCAATTCCTGTGAAGAAAGTAA 501  
Db 3174 CCAATTCCTGTGAAGAAAGTCA 3154

RESULT 4

US-09-731-231A-3  
; Sequence 3, Application US/09731231A  
; Patent No. US20020082189A1  
; GENERAL INFORMATION:  
; APPLICANT: MERKULOV et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL001007  
; CURRENT APPLICATION NUMBER: US/09/731,231A  
; CURRENT FILING DATE: 2000-12-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 326014  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(326014)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-731-231A-3

Query Match 94.4%; Score 472.8; DB 10; Length 326014;  
Best Local Similarity 96.8%; Pred. No. 2.7e-120;  
Matches 483; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 2 TTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTGAATGGTATTGCTGG 61  
Db 31241 TTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTGAATGGTATTGCTGG 31300  
QY 62 GTTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTGA 121  
Db 31301 GTTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTGA 31360  
QY 122 ATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTAG 181  
Db 31361 ATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTAG 31420

QY 182 CCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCTC 241  
Db 31421 CCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCTC 31480  
QY 242 AGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGAGGCTCTGTCT 301  
Db 31481 AGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGAGGCTCTGTCT 31540  
QY 302 GTTCCATTTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTTACTGTAG 361  
Db 31541 GTTCCATTTGTTGATATCTCTCTTTTGGTACCAGTACCAGTGTGTTGGTTTACTGTAG 31600  
QY 362 CTTGTAGTGTAGTTTGAAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTAG 421  
Db 31601 CTTGTAGTATAGTTTGAAGTCAAGTATGATGCTCCAGCTTTGTTCTTTTGGCTTAG 31660  
QY 422 GATTGACTGGCGATGTGGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTTTCTC 481  
Db 31661 GATTGACTGGCGATGTGGGCTCTTTTGGTTCACATGAACATTTAAAGTAGTTTCTC 31720  
QY 482 CAATTCCTGTGAAGAAAGTAA 501  
Db 31721 CAATTCCTGTGAAGAAAGTCA 31740

RESULT 5

US-09-777-921A-3/C  
; Sequence 3, Application US/09777921A  
; Patent No. US20020115136A1  
; GENERAL INFORMATION:  
; APPLICANT: MERKULOV et al.  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001103  
; CURRENT APPLICATION NUMBER: US/09/777,921A  
; CURRENT FILING DATE: 2002-02-07  
; NUMBER OF SEQ ID NOS: 126  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 69327  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(69327)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-921A-3

Query Match 94.3%; Score 472.2; DB 10; Length 69327;  
Best Local Similarity 96.4%; Pred. No. 1.9e-120;  
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
QY 1 CTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTGAATGGTATTGCTGG 60  
Db 40174 CTTTGGTGTATAGCATGAAGTCCCTTGCCATGCTATGCTGAATGGTATTGCTGG 40115  
QY 61 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTG 120  
Db 40114 GGTTCCTCTAGGGATTTTATGGTTTATAGTCTTAACATTTAAGTCTTTAATCCATCTG 40055  
QY 121 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180  
Db 40054 AATTAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 39995  
QY 181 GCCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCT 240  
Db 39994 GCCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATTCCTGCTTTTCT 39935  
QY 241 CAGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGAGGCTCTGTCT 300  
Db 39934 CAGTTTGTCAAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTAGAGGCTCTGTCT 39875

QY	301	TGTTCCATTGGTTGATATCTCTGTTTGGTACCAGTACCATGTTCTTTGGTTACTGTAG	360
Db	39874	TGTTCCATTGGTCTATATCTCTGTTTGGTACCAGTACCATGCTGGTTGGTTACTGTAG	39815
QY	361	CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTGTCTTTTGGCCTTA	420
Db	39814	CCTTGTAGTATAGTTTGAAGTCAGGTAGCATGATGCTCCAGCTTGTCTTTTGGCCTTA	39755
QY	421	GGATTGACTGGCGATGTGGGCTCTTTTGGTTCCACATGAACATTTAAAGTAGTTTTT	480
Db	39754	GGATTGACTTGGCAATGCGGCTCTTTTGGTTCCATATGAACATTTAAAGTAGTTTTT	39695
QY	481	CCAAATCTCTGAAGAAGTAA	501
Db	39694	CCAAATCTCTGAAGAAGTCA	39674
RESULT 6			
US-09-962-436-281/c			
; Sequence 281, Application US/09962436			
; Patent No. US20020081301A1			
; GENERAL INFORMATION:			
; APPLICANT: Soppet, Daniel			
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using			
; TITLE OF INVENTION: Sets			
; FILE REFERENCE: 689290-75			
; CURRENT APPLICATION NUMBER: US/09/962,436			
; CURRENT FILING DATE: 2001-09-25			
; PRIOR APPLICATION NUMBER: US/60/235,082			
; PRIOR FILING DATE: 2000-09-25			
; PRIOR APPLICATION NUMBER: US/60/234,924			
; PRIOR FILING DATE: 2000-09-25			
; NUMBER OF SEQ ID NOS: 568			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 281			
; LENGTH: 167343			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-962-436-281			
Query Match 94.3%; Score 472.2; DB 10; Length 167343;			
Best Local Similarity 96.4%; Pred. No. 2.9e-120;			
Matches 483; Conservative 0; Mismatches 18; Indels 0; Gaps 0;			
QY	1	CTTTTGGTGTTTAGACATGAAGTCCTTGCCTATGCTTGCCTGAATGGTATTGCTG	60
Db	59330	CTTTTGGTGTTTAGACATGAAGTCCTTGCCTATGCTTGAAGGTAATGCCTA	59271
QY	61	GGTTTCTTCTAGGGAATTTATGGTTTATAGGTCATAACATTTAAGTCTTTAATCCATCTTG	120
Db	59270	GGTTTCTTCTAGGTTTATGGTTTATAGGTCATAACGTTTAAGTCTTTAATCCATCTTG	59211
QY	121	AATTAATTTTGTATAAAGTGTAAGGAAGGATCCAGTTCAGCTTTCTACATAGGGCTA	180
Db	59210	AATTAATTTTGTATAAAGTGTAAGGAAGGATCCAGTTCAGCTTTCTACATAGGGCTA	59151
QY	181	GCCAGTTTCTCAGCACCATTTTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTTCT	240
Db	59150	GCCAGTTTCTCAGCACCATTTTATTAATAGGGAATCCTTTCCCATTCGTTGTTTTTCT	59091
QY	241	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGGCGCATTAATTCTCAGGGCTCTGTTTC	300
Db	59090	CAGGTTTGTCAAAGATCAGATAGTTGTAGATATGGCGGTATTATTCTCAGGGCTCTGTTTC	59031
QY	301	TGTTCCATTGGTTGATATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG	360
Db	59030	TGTTCCATTGGTCTATATCTCTGTTTGGTACCAGTACCATGCTGTTTGGTTACTGTAG	58971
QY	361	CCTTGTAGTGTAGTTTGAAGTCAGGTAGCATGATGCCCTCAGCTTGTCTTTTGGCCTTA	420
Db	58970	CCTTGTAGTATAGTTTGAAGTCAGGTAGCATGATGCCCTCAGCTTGTCTTTTGGCCTTA	58911

RESULT 8

US-10-003-806-6  
; Sequence 6, Application US/10003806  
; Patent No. US20020119929A1  
; GENERAL INFORMATION:  
; APPLICANT: Bishop, Colin E.  
; APPLICANT: Agoulnik, Alexander I.  
; APPLICANT: Zhu, Qichao  
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY  
; FILE REFERENCE: P02066US1/10024824  
; CURRENT APPLICATION NUMBER: US/10/003,806  
; PRIOR FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 60/245,872  
; PRIOR FILING DATE: 2000-11-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6  
; LENGTH: 180557  
; TYPE: DNA  
; ORGANISM: Human  
US-10-003-806-6

Query Match 93.6%; Score 469; DB 12; Length 180557;  
Best Local Similarity 96.0%; Pred. No. 2.3e-119;  
Matches 481; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
QY 1 CTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCCTATGCTCCTGAATGGTATGCTG 60  
Db CTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCCTATGCTCCTGAATGGTATGCTG 70370  
QY 61 GGTTCCTCTAGGAGTTTATGGTTTAAAGTCTAACATTTAAGTCTTAAATCCCATCTTG 120  
Db GGTTCCTCTAGGAGTTTATGGTTTAAAGTCTAACATTTAAGTCTTAAATCCCATCTTG 70430  
QY 121 AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGGCTA 180  
Db AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGGCTA 70490  
QY 181 GCCAGTTTCTCAGACACCAATTTAATAATAGGAATCCCTTCCCATGCTTGTGTTTCT 240  
Db GCCAGTTTCTCAGACACCAATTTAATAATAGGAATCCCTTCCCATGCTTGTGTTTCT 70490  
QY 241 CAGTTTGTCAAGATCAGATAGTTGATGCGGCAATTTTCTGAGGCTCTGTTCT 300  
Db CAGTTTGTCAAGATCAGATAGTTGATGCGGCAATTTTCTGAGGCTCTGTTCT 70550  
QY 301 TGTTCATTTGATATCTCTGTTTGTACACAGTACCATGTTGTTTGGTACTGTAG 360  
Db TGTTCATTTGATATCTCTGTTTGTACACAGTACCATGTTGTTTGGTACTGTAG 70610  
QY 361 CCTTGTAGTGTAGTTTGAAGTCAAGTATGATGCGGCAATTTTCTGAGGCTCTGTTCT 420  
Db CCTTGTAGTGTAGTTTGAAGTCAAGTATGATGCGGCAATTTTCTGAGGCTCTGTTCT 70550  
QY 421 GGATTCACATGCGGATGCTGCTCTTTTGGTCCACATGAACCTTAAAGTAGTTTTT 480  
Db GGATTCACATGCGGATGCTGCTCTTTTGGTCCACATGAACCTTAAAGTAGTTTTT 70670  
QY 481 CCAATTTCTGTGAAGAAAGTAA 501  
Db CCAATTTCTGTGAAGAAAGTCA 70811

RESULT 9

US-10-003-806-9  
; Sequence 9, Application US/10003806  
; Patent No. US20020119929A1  
; GENERAL INFORMATION:  
; APPLICANT: Bishop, Colin E.  
; APPLICANT: Agoulnik, Alexander I.  
; APPLICANT: Zhu, Qichao

; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY  
; FILE REFERENCE: P02066US1/10024824  
; CURRENT APPLICATION NUMBER: US/10/003,806  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 60/245,872  
; PRIOR FILING DATE: 2000-11-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 9  
; LENGTH: 180557  
; TYPE: DNA  
; ORGANISM: Human  
US-10-003-806-9

Query Match 93.6%; Score 469; DB 12; Length 180557;  
Best Local Similarity 96.0%; Pred. No. 2.3e-119;  
Matches 481; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
QY 1 CTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCCTATGCTCCTGAATGGTATGCTG 60  
Db CTTTGGTCTTTAGACATGAAGTCTTGCCTCCCATGCCTATGCTCCTGAATGGTATGCTG 70370  
QY 61 GGTTCCTCTAGGAGTTTATGGTTTAAAGTCTAACATTTAAGTCTTAAATCCCATCTTG 120  
Db GGTTCCTCTAGGAGTTTATGGTTTAAAGTCTAACATTTAAGTCTTAAATCCCATCTTG 70430  
QY 121 AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGGCTA 180  
Db AATTAATTTTGTATAAGTGTGAAGGAAGGATCCAGTTTTCAGCTTTTACATAGGGCTA 70490  
QY 181 GCCAGTTTCTCAGACACCAATTTAATAATAGGAATCCCTTCCCATGCTTGTGTTTCT 240  
Db GCCAGTTTCTCAGACACCAATTTAATAATAGGAATCCCTTCCCATGCTTGTGTTTCT 70550  
QY 241 CAGTTTGTCAAGATCAGATAGTTGATGCGGCAATTTTCTGAGGCTCTGTTCT 300  
Db CAGTTTGTCAAGATCAGATAGTTGATGCGGCAATTTTCTGAGGCTCTGTTCT 70610  
QY 301 TGTTCATTTGATATCTCTGTTTGTACACAGTACCATGTTGTTTGGTACTGTAG 360  
Db TGTTCATTTGATATCTCTGTTTGTACACAGTACCATGTTGTTTGGTACTGTAG 70670  
QY 361 CCTTGTAGTGTAGTTTGAAGTCAAGTATGATGCGGCAATTTTCTGAGGCTCTGTTCT 420  
Db CCTTGTAGTGTAGTTTGAAGTCAAGTATGATGCGGCAATTTTCTGAGGCTCTGTTCT 70730  
QY 421 GGATTCACATGCGGATGCTGCTCTTTTGGTCCACATGAACCTTAAAGTAGTTTTT 480  
Db GGATTCACATGCGGATGCTGCTCTTTTGGTCCACATGAACCTTAAAGTAGTTTTT 70790  
QY 481 CCAATTTCTGTGAAGAAAGTAA 501  
Db CCAATTTCTGTGAAGAAAGTCA 70811

RESULT 10

US-09-764-891-8497/C  
; Sequence 8497, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8497  
; LENGTH: 6684  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-8497

Query Match 93.3%; Score 467.4; DB 9; Length 6684;  
Best Local Similarity 95.8%; Pred. No. 1.3e-119;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTG 60  
DB CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTA 13236

QY 61 GGTTCCTCTAGGATTTTATGTTTAAAGTCTAAACATTTAAAGTCTTTAAATCCATCTTG 120  
DB CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTA 5560

QY 121 AATTAATTTTGTATAAGGCTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
DB GGTTCCTCTAGGATTTTATGTTTAAAGTCTAAACATTTAAAGTCTTTAAATCCATCTTG 5500

QY 181 GCCAGTTTCTCAGACACATTTATTAATAGGGAATCCTTCCCATGCTTTGTTTCT 240  
DB GCCAGTTTCTCAGACACATTTATTAATAGGGAATCCTTCCCATGCTTTGTTTCT 5380

QY 241 CAGGTTCTCAAGATCAGATAGTTGTAGATATGCGCATTTATTCGAGGCTCTGTC 300  
DB CAGGTTCTCAAGATCAGATAGTTGTAGATATGCGCATTTATTCGAGGCTCTGTC 5320

QY 301 TGTTCATTTGATATCTCTGTTTGGTACCAGTACCAGTCTGTTTGGTACTGTC 360  
DB TGTTCATTTGATATCTCTGTTTGGTACCAGTACCAGTCTGTTTGGTACTGTC 5260

QY 361 CTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTTGTTTGGCTTA 420  
DB CTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTTGTTTGGCTTA 5200

QY 421 GGATTGACTGCGCATGCGGCTCTTTTGGTCCATGAACCTTTAAAGTACTGTTT 480  
DB GGATTGACTGCGCATGCGGCTCTTTTGGTCCATGAACCTTTAAAGTACTGTTT 5140

QY 481 CCAATTCGTGAAGAAAGTAA 501  
DB CCAATTCGTGAAGAAAGTCA 5119

RESULT 11  
US-10-118-984-44  
; Sequence 44, Application US/10118984  
; Publication No. US20020197693A1  
; GENERAL INFORMATION:  
; APPLICANT: Bertin, John  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY  
; FILE REFERENCE: 07334/118001  
; CURRENT APPLICATION NUMBER: US/10/118,984  
; PRIOR FILING DATE: 2002-04-09  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/245,281  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-05  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/207,359  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/099,041  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/019,942  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 32042  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-118-984-44

Query Match 93.3%; Score 467.4; DB 9; Length 32042;  
Best Local Similarity 95.8%; Pred. No. 2.8e-119;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTG 60  
DB CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTA 13236

QY 61 GGTTCCTCTAGGATTTTATGTTTAAAGTCTAAACATTTAAAGTCTTTAAATCCATCTTG 120  
DB CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTG 13296

QY 121 AATTAATTTTGTATAAGGCTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
DB AATTAATTTTGTATAAGGCTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGGCTA 13356

QY 181 GCCAGTTTCTCAGACACATTTATTAATAGGGAATCCTTCCCATGCTTTGTTTCT 240  
DB GCCAGTTTCTCAGACACATTTATTAATAGGGAATCCTTCCCATGCTTTGTTTCT 13416

QY 241 CAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGCATTTATTCGAGGCTCTGTC 300  
DB CAGGTTTGTCAAGATCAGATAGTTGTAGATATGCGCATTTATTCGAGGCTCTGTC 13476

QY 301 TGTTCATTTGATATCTCTGTTTGGTACCAGTACCAGTCTGTTTGGTACTGTC 360  
DB TGTTCATTTGATATCTCTGTTTGGTACCAGTACCAGTCTGTTTGGTACTGTC 13536

QY 361 CTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTTGTTTGGCTTA 420  
DB CTTGTAGTGTAGTTGAAGTCAGTACGATGATGCTCCAGCTTTGTTTGGCTTA 13596

QY 421 GGATTGACTGCGCATGCGGCTCTTTTGGTCCATGAACCTTTAAAGTACTGTTT 480  
DB GGATTGACTGCGCATGCGGCTCTTTTGGTCCATGAACCTTTAAAGTACTGTTT 13656

QY 481 CCAATTCGTGAAGAAAGTAA 501  
DB CCAATTCGTGAAGAAAGTCA 13677

RESULT 12  
US-10-295-981-63  
; Sequence 63, Application US/10295981  
; Publication No. US20030120055A1  
; GENERAL INFORMATION:  
; APPLICANT: Bertin, John  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES TH  
; FILE REFERENCE: 07334-124001  
; CURRENT APPLICATION NUMBER: US/10/295,981  
; CURRENT FILING DATE: 2002-11-15  
; PRIOR APPLICATION NUMBER: US/09/340,620  
; PRIOR FILING DATE: 1999-06-28  
; PRIOR APPLICATION NUMBER: US 09/245,281  
; PRIOR FILING DATE: 1999-02-05  
; PRIOR APPLICATION NUMBER: US 09/207,359  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: US 09/099,041  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: US 09/019,942  
; PRIOR FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63  
; LENGTH: 32042  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-295-981-63

Query Match 93.3%; Score 467.4; DB 9; Length 32042;  
Best Local Similarity 95.8%; Pred. No. 2.8e-119;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTG 60  
DB CTTTGGTCTTTAGACATGAAGTCCCTTGCCCATGCCCTATGCTCCTGAATGTTATGCGCTA 13236



QY 61 GGTTCCTCTAGGAGTTTATGGTTTATAGGCTTAACATTTAAGCTTTTAAATCCATCTTG 120  
DB 13237 GGTTCCTCTAGGAGTTTATGGTTTATAGGCTTTAAACGTTTAAATCTTTAAATCCATCTTG 13296  
QY 121 AATTAATTTCTATAAGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180  
DB 13297 AATTGATTTTCTATAAGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356  
QY 181 GCCAGTTTCTCAGCACCATTTTAAATAGGGAATCCTTTCCCATTTGCTTTGTTTTCT 240  
DB 13357 GCCAGTTTCCAGCACCATTTTAAATAGGGAATCCTTTCCCATTTGCTTTGTTTTCT 13416  
QY 241 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 300  
DB 13417 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 13476  
QY 301 TCTTCCATTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 360  
DB 13477 TCTTCCATTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 13536  
QY 361 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 420  
DB 13537 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 13596  
QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 480  
DB 13597 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 13656  
QY 481 CCAATTCCTGGAAGAAAGTAA 501  
DB 13657 CCAATTCCTGGAAGAAAGTCA 13677

RESULT 13  
US-09-728-721-63  
; Sequence 63, Application US/09728721  
; Patent No. US20020061845A1  
; GENERAL INFORMATION:  
; APPLICANT: Bertin, John  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE  
; CURRENT FILING DATE: 2000-12-01  
; CURRENT APPLICATION NUMBER: US/09/728,721  
; PRIOR FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: 09/340,620  
; PRIOR FILING DATE: 1999-06-28  
; PRIOR APPLICATION NUMBER: US 09/207,359  
; PRIOR FILING DATE: 1998-12-08  
; PRIOR APPLICATION NUMBER: US 09/099,041  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: US 09/019,942  
; PRIOR FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 63  
; LENGTH: 32042  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-728-721-63

Query Match 93.3%; Score 467.4; DB 10; Length 32042;  
Best Local Similarity 95.8%; Pred. No. 2.8e-119;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 CTTTGGTCTTTTACACATGAAGTCTTGCCTATGCTCCATGCTTAAAGCTTTTAAATCCATCTTG 60  
DB 13177 CTTTGGTCTTTGACATGAAGTCTTGCCTATGCTCCATGCTTAAAGCTTTTAAATCCATCTTG 13296  
QY 61 GGTTCCTCTAGGAGTTTATGGTTTATAGGCTTAACATTTAAGCTTTTAAATCCATCTTG 120  
DB 13237 GGTTCCTCTAGGAGTTTATGGTTTATAGGCTTTAAACGTTTAAATCTTTAAATCCATCTTG 13296  
QY 121 AATTAATTTTCTATAAGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 180  
DB 13297 AATTGATTTTCTATAAGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356

DB 13297 AATTGATTTTCTATAAGTGTAAGGAAGGGATCCAGTTTTCAGCTTTCTACATAGGGCTA 13356  
QY 181 GCCAGTTTCTCAGCACCATTTTAAATAGGGAATCCTTTCCCATTTGCTTTGTTTTCT 240  
DB 13357 GCCAGTTTCCAGCACCATTTTAAATAGGGAATCCTTTCCCATTTGCTTTGTTTTCT 13416  
QY 241 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 300  
DB 13417 CAGGTTTGCAAGATCAGATAGTTGTAGATATGCGGCATTAATTTCTGAGGGCTCTGTTC 13476  
QY 301 TCTTCCATTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 360  
DB 13477 TCTTCCATTTGATATCTCTGTTTGGTACCACTACCATGTTGTTTGGTTACTGTAG 13536  
QY 361 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 420  
DB 13537 CCTTGTAGTGTAGTTGAAGTCAGGTAGCATGCTCCAGCTTTGTTTCTTTTGGCTTA 13596  
QY 421 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 480  
DB 13597 GGATTGACTTGGCGATGGGCTCTTTTGGTCCACATGAACCTTTAAAGTAGTTTTT 13656  
QY 481 CCAATTCCTGGAAGAAAGTAA 501  
DB 13657 CCAATTCCTGGAAGAAAGTCA 13677

RESULT 14  
US-10-006-883A-15  
; Sequence 15, Application US/10006883A  
; Publication No. US20030119767A1  
; GENERAL INFORMATION:  
; APPLICANT: Kenneth W. Dobie  
; TITLE OF INVENTION: ANTISENSE MODULATION OF NOD1 EXPRESSION  
; FILE REFERENCE: RTS-0337  
; CURRENT APPLICATION NUMBER: US/10/006,883A  
; CURRENT FILING DATE: 2001-12-05  
; NUMBER OF SEQ ID NOS: 96  
; SEQ ID NO 15  
; LENGTH: 34001  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: (425)...(524)  
; LOCATION: (425)...(524)  
; OTHER INFORMATION: unknown  
; FEATURE:  
; NAME/KEY: Intron:exon junction  
; LOCATION: (10779)...(10780)  
; OTHER INFORMATION: Intron 1:exon 2  
; FEATURE:  
; NAME/KEY: Intron:exon junction  
; LOCATION: (9546)...(9547)  
; OTHER INFORMATION: Intron 7b:exon 8b  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (10878)...(10961)  
; OTHER INFORMATION: exon 8  
; FEATURE:  
; NAME/KEY: Intron:exon junction  
; LOCATION: (21846)...(21847)  
; OTHER INFORMATION: Intron 10:exon 11  
; FEATURE:  
; NAME/KEY: exon:intron junction  
; LOCATION: (21930)...(21931)  
; OTHER INFORMATION: exon 11:intron 11  
; FEATURE:  
; NAME/KEY: Intron  
; LOCATION: (21931)...(24748)  
; OTHER INFORMATION: Intron 11  
; FEATURE:  
; NAME/KEY: exon:intron junction  
; LOCATION: (29424)...(29425)

; OTHER INFORMATION: exon 14:intron 14  
; FEATURE:  
; NAME/KEY: Intron  
; LOCATION: (29425)...(32217)  
; OTHER INFORMATION: intron 14  
; FEATURE:  
; NAME/KEY: Intron:exon junction  
; LOCATION: (32217)...(32218)  
; OTHER INFORMATION: intron 14:exon 15  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (32218)...(33394)  
; OTHER INFORMATION: exon 15  
US-10-067-514-15

Query Match 93.3%; Score 467.4; DB 9; Length 34001;  
Best Local Similarity 95.8%; Pred. No. 2.9e-119;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGTAATGCTATGCGCTG 60  
Db CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGTAATGCTATGCGCTG 60  
Qy 61 GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 120  
Db GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 120  
Qy 121 AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
Db AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
Qy 13759 GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 13818  
Db GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 13818  
Qy 13819 AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
Db AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
Qy 181 GCCAGTTTTCACAGACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 240  
Db GCCAGTTTTCACAGACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 240  
Qy 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCAATTTATTTCTGAGGGCTCTGTT 300  
Db CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCAATTTATTTCTGAGGGCTCTGTT 300  
Qy 301 TGTTCATTTGTTGATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360  
Db TGTTCATTTGTTGATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360  
Qy 361 CCTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420  
Db CCTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420  
Qy 421 GGATTTGACTTGGCGATGTTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480  
Db GGATTTGACTTGGCGATGTTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480  
Qy 481 CCAATTTCTGTGAAGAAATGTA 501  
Db CCAATTTCTGTGAAGAAATGTA 501

RESULT 15  
US-10-067-514-1/c  
; Sequence 1, Application US/10067514  
; Publication No. US20030054531A1  
; GENERAL INFORMATION:  
; APPLICANT: GretaSdottir, Solveig  
; APPLICANT: Jonsdottir, Sif  
; APPLICANT: Reynisdottir, Sigríður Th.  
; TITLE OF INVENTION: HUMAN STROKE GENE  
; FILE REFERENCE: 2345-2010-003  
; CURRENT APPLICATION NUMBER: US/10/067,514  
; PRIOR FILING DATE: 2002-02-04  
; PRIOR APPLICATION NUMBER: US 09/811/352  
; PRIOR FILING DATE: 2001-03-19  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1

; LENGTH: 1691139  
; TYPE: DNA  
; ORGANISM: Human  
US-10-067-514-1

Query Match 93.3%; Score 467.4; DB 9; Length 1691139;  
Best Local Similarity 95.8%; Pred. No. 1.9e-118;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGTAATGCTATGCGCTG 60  
Db CTTTGGTGTATAGACATGAAGTCCTTGCCCATGCCCTATGCTCGTAATGCTATGCGCTA 59202  
Qy 61 GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 120  
Db GGTTCCTCTAGGGAATTTATGTTTATGCTCTAAGCTTTAAGCTTTAATCCATCTTG 59142  
Qy 121 AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 180  
Db AATTAATTTTGTATAGGCTAAGGAGGATCCAGTTTCAGCTTTTACATAGGGCTA 59082  
Qy 181 GCCAGTTTTCACAGACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 240  
Db GCCAGTTTTCACAGACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTTCT 59022  
Qy 241 CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCAATTTATTTCTGAGGGCTCTGTT 300  
Db CAGGTTTCTCAAGATCAGATAGTTGTAGATATCGGCAATTTATTTCTGAGGGCTCTGTT 59062  
Qy 301 TGTTCATTTGTTGATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 360  
Db TGTTCATTTGTTGATCTCTGTTTGGTACCAGTACCATGTTGTTTGGTTACTGTAG 58902  
Qy 361 CCTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 420  
Db CCTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGGCTTA 58842  
Qy 421 GGATTTGACTTGGCGATGTTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480  
Db GGATTTGACTTGGCGATGTTGGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 58782  
Qy 481 CCAATTTCTGTGAAGAAATGTA 501  
Db CCAATTTCTGTGAAGAAATGTA 58761

Search completed: July 4, 2003, 19:36:19  
Job time : 118.075 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:16:37 ; Search time 35.0233 Seconds  
(without alignments)  
4386.938 Million cell updates/sec

Title: US-10-083-853B-2\_COPY\_29421\_29921  
Perfect score: 501  
Sequence: 1 ctttgggttttagacatg.....caattctgtgaagaagtaa 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents\_NA: \*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467.4	93.3	32042	4	US-09-245-281-44
2	461.2	92.1	40328	3	US-08-742-185-102
3	452.4	90.3	18596	4	US-09-318-448-11
4	442.6	88.3	168575	4	US-09-426-290-1
5	440.4	87.9	10607	1	US-08-078-090-3
6	410.2	81.9	50000	4	US-09-146-053-4
7	319.2	63.7	10754	2	US-08-966-958-1
8	319.2	63.7	10754	2	US-09-215-817-1
9	319.2	63.7	10754	4	US-09-342-353-1
10	266.2	53.1	152331	3	US-09-128-155-16
11	266.2	53.1	176373	3	US-09-128-155-17
12	248.4	49.6	80246	4	US-09-078-294-4
13	248.4	49.6	80595	4	US-09-078-294-3
14	246.8	49.3	50000	4	US-09-146-053-3
15	246.2	49.1	38844	4	US-09-734-675-3
16	229.4	45.8	168575	4	US-09-426-290-1
17	187.2	37.4	22067	4	US-09-820-001-3
18	185	36.9	112132	4	US-09-741-150-3
19	184	36.7	98844	4	US-09-791-211-10
20	181	36.1	29604	3	US-08-781-891-207
21	174	34.7	624	4	US-09-385-982-209
22	165.2	33.0	1200	4	US-09-018-584A-37
23	161	32.1	10409	3	US-08-772-440-33
24	150.6	30.1	6367	1	US-08-470-299-1
25	150.6	30.1	6367	3	US-08-776-511-3
26	150.6	30.1	6926	1	US-08-470-299-2
27	144.8	28.9	55827	4	US-09-813-133A-3

C	28	133.2	26.6	3450	4	US-09-462-561B-10	Sequence 10, Appl
C	29	124.6	24.9	112132	4	US-09-741-150-3	Sequence 3, Appl
C	30	120.6	24.1	3430	4	US-09-462-561B-9	Sequence 9, Appl
C	31	119.8	23.9	4693	4	US-09-462-561B-8	Sequence 8, Appl
C	32	119.8	23.9	38564	4	US-09-734-673-3	Sequence 3, Appl
C	33	110.6	22.1	36159	4	US-09-749-588-3	Sequence 3, Appl
C	34	103.6	20.7	29604	3	US-08-781-891-207	Sequence 207, App
C	35	98	19.6	694	4	US-09-605-785-604	Sequence 604, App
C	36	98	19.6	694	4	US-09-605-785-665	Sequence 665, App
C	37	95.8	19.1	90050	4	US-09-245-041-5	Sequence 5, Appl
C	38	95.6	19.1	176373	3	US-09-128-155-17	Sequence 17, Appl
C	39	95.2	19.0	43795	3	US-08-742-185-101	Sequence 101, App
C	40	90.4	18.0	4344	4	US-09-462-561B-11	Sequence 11, Appl
C	41	90.2	18.0	705	4	US-09-605-785-666	Sequence 666, App
C	42	89.4	17.8	11811	4	US-09-078-294-7	Sequence 7, Appl
C	43	88.6	17.7	246240	2	US-08-724-394A-20	Sequence 20, Appl
C	44	88.6	17.7	246240	2	US-08-724-394A-21	Sequence 21, Appl
C	45	88.6	17.7	246240	2	US-08-724-394A-22	Sequence 22, Appl

## ALIGNMENTS

RESULT 1  
US-09-245-281-44  
; Sequence 44, Application US/09245281  
; Patent No. 6369196  
; GENERAL INFORMATION:  
; APPLICANT: Bertin, John  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY  
; FILE REFERENCE: 07334/118001  
; CURRENT APPLICATION NUMBER: US/09/245,281  
; CURRENT FILING DATE: 1999-02-05  
; EARLIER APPLICATION NUMBER: US 09/207,359  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: US 09/099,041  
; EARLIER FILING DATE: 1998-06-17  
; EARLIER APPLICATION NUMBER: US 09/019,942  
; EARLIER FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 32042  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-245-281-44

Query Match 93.3%; Score 467.4; DB 4; Length 32042;  
Best Local Similarity 95.8%; Pred.No. 6e-129;  
Matches 480; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTTTGGTGTTTAGACATGAAGTCTCTGCCCATGCTATGCTCTGAATGGTAAATGCTG 60  
DB 13177 CTTTGGTGTTTGGACATGAAGTCTCTGCCCATGCTATGCTCTGAATGGTAAATGCTG 13236  
QY 61 GGTTCCTCTAGGGATTTATGGTTTATAGTCTAACTTAAGTCTTTAACTCACTTCTG 120  
DB 13237 GGTTCCTCTAGGGATTTATGGTTTATAGTCTAACTTAAGTCTTTAACTCACTTCTG 13296  
QY 121 AATTAATTTTCTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTCACATAGGCTA 180  
DB 13297 AATTGATTTTGTATAAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTCACATAGGCTA 13356  
QY 181 GCAGTTTCTCAGCACCATTTATTAATAGGAATCCCTTCCCATTTCTGCTGTTTTCT 240  
DB 13357 GCAGTTTCTCAGCACCATTTATTAATAGGAATCCCTTCCCATTTCTGCTGTTTTCT 13416  
QY 241 CAGGTTTGTCAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 300  
DB 13417 CAGGTTTGTCAAGATCAGATAGTCTAGATATGCGGCATTTATTTCTGAGGCTCTGTC 13476  
QY 301 TGTTCATTGGTTGATATCTCTGTTTGTACCAGTACCATGTTGTTTGTGTTACTGTAG 360

Db 13477 TGTTCATGATCTATCTCTGTTTGGTACCAGTACCAGCTGTTTGGTTACTGTAG 13536  
Qy 361 CCTTGTAGTGTAGTTGAAGTCAGGTACGATCCCTCCAGCTTTGTTCTTTTGGCTTA 420  
Db 13537 CCTGTAGTATAGTTGAAGTCAGGTACGATCCCTCCAGCTTTGTTCTTTTGGCTTA 13596  
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480  
Db 13597 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCATATGAACCTTTAAAGTAGTTTTT 13656  
Qy 481 CCAATTCGTGAAGAAAGTAA 501  
Db 13657 CCAATTCGTGAAGAAAGTCA 13677

## RESULT 2

US-08-742-185-102  
; Sequence 102, Application US/08742185  
; Patent No. 6020476  
; GENERAL INFORMATION:  
; APPLICANT: Page, David C.  
; APPLICANT: Reijo, Renee  
; APPLICANT: Saxena, Richa  
; APPLICANT: Hawkins, Trevor  
; APPLICANT: Reeve, Mary Pat  
; TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02173

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/742,185  
; FILING DATE: 30-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/690,734  
; FILING DATE: 31-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/310,429  
; FILING DATE: 22-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WHI94-07A2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 102:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40328 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-742-185-102

Query Match 92.1%; Score 461.2; DB 3; Length 40328;  
Best Local Similarity 95.4%; Pred. No. 4.6e-127;  
Matches 475; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 CTTTGTGTTTGTAGACATGAAGTCCCTTGGCCATGCCTATGCTCTGATGCTATTTGCTG 60  
Db 17396 CTTTGTGTTTGTAGACATGAAGTCCCTGCGCTATGCTCTGATGCTATTAATGCTA 17455

Qy 61 GGTTCCTCTTAGGATTTTATGTTTGTAGCTCTAAACATTTAAGTCTTTTAATCCATCTTG 120  
Db 17456 GGTTCCTCTTAGGATTTTATGTTTGTAGCTCTAAACATTTAAGTCTTTTAATCCATCTTG 17515  
Qy 121 AATTAATTTTGTATAGGTCTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 180  
Db 17516 AATTAATTTTGTATAGGTCTAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGGCTA 17575  
Qy 181 GCCAGTTTCTCAGCACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTCT 240  
Db 17576 GCCAGTTTCTCAGCACCATTTAATAAGGGAATCCTTTCCCATGCTTTGTTTCT 17635  
Qy 241 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGCAATTTATTTGAGGGCTCTGTTC 300  
Db 17636 CAGGTTTGTCAAAGATCAGATAGTTAGATATCGGCGTTATTTCTGAGGGCTCTGTTC 17695  
Qy 301 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 360  
Db 17696 TGTTCATGTTGATATCTCTGTTTGGTACCAGTACCAGTGTGTTTGGTTACTGTAG 17755  
Qy 361 CTTGTAGTGTAGTTTGAAGTCAGTAGCATGATGATCCCTCCAGCTTTGTTCTTTGGCTTA 420  
Db 17756 CTTGTAGTGTAGTTTGAAGTCAGTAGGTGTGATGCCCTCCAGCTTTGTTCTTTGGCTTA 17815  
Qy 421 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCACATGAACCTTTAAAGTAGTTTTT 480  
Db 17816 GGATTGACTTGGCGATGCGGCTCTTTTGGTTCCATGTCACTTTAAAGTAGTTTTT 17875  
Qy 481 CCAATTCGTGAAGAAAG 498  
Db 17876 CCAATTCGTGAAGAAAG 17893

## RESULT 3

US-09-318-448-11  
; Sequence 11, Application US/09318448  
; Patent No. 6210950  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, William G.  
; APPLICANT: Stenroos, Edward S.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING  
; FILE REFERENCE: 601-1-057  
; CURRENT APPLICATION NUMBER: US/09/318,448  
; CURRENT FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 18596  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-318-448-11

Query Match 90.3%; Score 452.4; DB 4; Length 18596;  
Best Local Similarity 94.7%; Pred. No. 1.4e-124;  
Matches 468; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 8 TGTTCATGACATGAAGTCCCTTGGCCATGCCTATGCTCTGATGCTATTTGCTGGTTTC 67  
Db 6806 TGCATAGGCATGAAGTCCCTTGGCCATGCCTATGCTCTGATGCTATTTGCTGGTTTC 6865  
Qy 68 TTCTAGGATTTTATGTTTGTAGCTCTAAACATTTAAGTCTTTAATCCATCTTGAATTAAT 127  
Db 6866 TTCTAGGATTTTATGTTTGTAGCTCTAAACATTTAAGTCTTTAATCCATCTTGAATTAAT 6925  
Qy 128 TTTTGTATAAGGTCTAAGGAAGGATCCAGTTTTCAGCTTTCTACATAGGGCTACCGAGTT 187  
Db 6926 TTTTGTATAAGGTCTAAGGAAGGATCCAGTTTTCAGCTTTTACATATGGCTAGCCAGTT 6985  
Qy 188 TTCTCAGCACCATTTTAAATAGGGAATCCTTTCCCATGCTTTGTTTCTCAGGTTT 247  
Db 6986 TTCCAGCACCATTTTAAATAGGGAATCCTTTCCCATGCTTTGTTTCTCAGGTTT 7045

```
QY 248 GTCAAGATCAGATAGTTAGATATCGCGCATTTATTTCTGAGGCTCTCTCTGTTCCA 307
Db 7046 GTCAAGATCAGATAGTTAGATATCGCGCGTTATTTCTGAGGCTCTCTCTGTTCCA 7105
QY 308 TTGGTTCATATCTCTTTTGGTACCAAGTACCAATGTTGTTGGTACTAGCTTGTA 367
Db 7106 TTGATCTATGTTGTTTGGTACCAAGTACCAATGTTGTTGGTACTAGCTTGTA 7165
QY 368 GTGTAGTTTCAAGTACAGTACGATGATCGCTCCAGCTTTCTCTTTGGCTTAGGATTGA 427
Db 7166 GTATAGTTTCAAGTACAGTACGATGATCGCTCCAGCTTTCTCTTTGGCTTAGGATTGA 7225
QY 428 CTTGGCATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTCCCAATTC 487
Db 7226 CTTGGCATGCGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTTCCCAATTC 7285
QY 488 TGTGAAGAAAGTAA 501
Db 7286 TGTGAAGAAAGTAA 7299

RESULT 4
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berghind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426.290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1
; Query Match 88.3%; Score 442.6; DB 4; Length 168575;
; Best Local Similarity 94.08; Pred. No. 2.7e-121;
; Matches 471; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 1 CTTTGGTGTTTTAGACATGAAGTCCCTTGGCCATGCTATGCTCTGAATGTTGCTG 60
Db 159839 CTTTGGTGTTTTAGACATGAAGTCCCTTGGCCATGCTATGCTCTGAATGTTG 159780
QY 61 GGTTCCTCTAGGATTTTATGGTTTAGCTTAACATTTAGTCTTTAATCCATCTTG 120
Db 159779 GGTTCCTCTAGGATTTTATGGTTTAGCTTAACATTTAGTCTTTAATCCATCTTG 159721
QY 121 AATTAATTTTGTATAGGTGAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180
Db 159720 AGTTAATTTTGTATAGGTGAAGGAAGGATCCAGTTTCAGCTTTCTACATAGGCTA 159661
QY 181 GCCAGTTTCTCAGCACCACTTTTAAATAGGAATCCTTTCCCAATGCTGTTTCT 240
```

```
Db 159660 ACCAGTTTCCAGCACCATTTATTAATAGGAATCCTTTCCCAATGCTGTTTCT 159601
QY 241 CAGTTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTTTC 300
Db 159600 CAGTTTTGTCAAAGATCAGATAGTTGTAGATATCGGCATTTATTTCTGAGGCTCTGTTTC 159541
QY 301 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTGTTGTTGTTACTGTAG 360
Db 159540 TGTTCATTTGTTGATATCTCTGTTTGGTACCAAGTACCAATGTTGTTGTTGTTACTGTAG 159481
QY 361 CTTTGTAGTGTAGTTTGAAGTACAGTACGATGCTCCAGCTTTGTTCTTTTGGCTTA 420
Db 159480 CTTTGTAGTGTAGTTTGAAGTACAGTACGATGCTCCAGCTTTGTTCTTTTGGCTTA 159421
QY 421 GGATTTGATTTGGCGATCTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 480
Db 159420 GGATTTGATTTGGCGATCTGGGCTCTTTTGGTTCACATGAACCTTTAAAGTAGTTTTT 159361
QY 481 CCAATTTCTGTGAAGAAAGTAA 501
Db 159360 CCAATTTCTGTGAAGAAAGTAA 159340

RESULT 5
US-08-078-090-3
; Sequence 3, Application US/08078090
; Patent No. 5739407
; GENERAL INFORMATION:
; APPLICANT: BERGSTROEM, SVEN
; APPLICANT: HERNELL, OLLE
; APPLICANT: LOENNERDAL, BO
; APPLICANT: HJALMARSSON, KARIN
; APPLICANT: HANSSON, LENNART
; APPLICANT: TOERNELL, JAN
; APPLICANT: STROEMQUIST, MATS
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING
; TITLE OF INVENTION: IT AND USE THEREOF
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/078,090
; FILING DATE: 19930618
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK92/00236
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/DK91/00233
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28, 005
; REFERENCE/DOCKET NUMBER: BERGSTROM2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)628-5197
; TELEFAX: (202)737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10607 base pairs
; TYPE: nucleic acid
```

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845  
;; LOCATION: ..6886, 7991..8521, 9440..9445)  
US-08-078-090-3

Query Match 87.9%; Score 440.4; DB 1; Length 10607;  
Best Local Similarity 93.6%; Pred. No. 3.9e-121;  
Matches 470; Conservative 0; Mismatches 31; Indels 1; Gaps 1;  
QY 1 CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAATGGTATGCGCTG 60  
Db CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAATGGTATGCGCTA 3154  
QY 61 GGTTCCTCTAGGATTTTATGGTTTATAGTCTTAACATTAAGTCTTTAATCCATCTTG 120  
Db GGTTCCTCTAGGATTTTATGGTTTATAGTCTTAACATTAAGTCTTTAATCCATCTTG 3214  
QY 121 AATTAATTTTGTATAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180  
Db AATTAATTTTGTATAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 3274  
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTTCCCATGCTTTCTTTTCT 240  
Db GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTTCCCATGCTTTCTTTTCT 3334  
QY 241 CAGGTTGTCAAAAGATCAGATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTTCC 300  
Db CAGGTTGTCAAAAGATCAGATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTTCC 3394  
QY 301 TGTTCATGTTGATATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTTCC 359  
Db TGTTCATGTTGATATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTTCC 3454  
QY 360 GCCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 419  
Db GCCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 3514  
QY 420 AGGATTGACTTGGCGATGGGCTCTTTTGTGTTCCACATGAACCTTTAAAGTAGTTT 479  
Db AGGATTGACTTGGCGATGGGCTCTTTTGTGTTCCACATGAACCTTTAAAGTAGTTT 3574  
QY 480 TCAATTCGTGAAGAAAGTAA 501  
Db TCAATTCGTGAAGAAAGTAA 3596

RESULT 6  
US-09-146-053-4  
; Sequence 4, Application US/09146053A  
; Patent No. 6399349  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, James W.  
; APPLICANT: Sprinkle, Terry Joe Curtis  
; APPLICANT: Venema, Richard C.  
; TITLE OF INVENTION: Human Aminopectidase P Gene  
; FILE REFERENCE: MCG103  
; CURRENT APPLICATION NUMBER: US/09/146, 053A  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER FILING DATE: 1997-09-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 50000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-146-053-4

Query Match 81.9%; Score 410.2; DB 4; Length 50000;

Best Local Similarity 93.8%; Pred. No. 6.9e-112;  
Matches 427; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 1 CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAATGGTATGCGCTG 60  
Db CTTTGGTGTAGACATGAAGTCTTGGCCATGCTATGCTCCTGAATGGTATGCGCTA 27099  
QY 61 GGTTCCTCTAGGATTTTATGGTTTATAGTCTTAACATTAAGTCTTTAATCCATCTTG 120  
Db GGTTCCTCTAGGATTTTATGGTTTATAGTCTTAACATTAAGTCTTTAATCCATCTTG 27159  
QY 121 AATTAATTTTGTATAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 180  
Db AATTAATTTTGTATAGTGTAAAGGAGGATCCAGTTTCAGCTTTCTACATAGGCTA 27219  
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTTCCCATGCTTTCTTTTCT 240  
Db GCCAGTTTCTCAGCACCATTATTAATAGGAATCCCTTTCCCATGCTTTCTTTTCT 27279  
QY 241 CAGGTTGTCAAAAGATCAGATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTGTT 300  
Db CAGGTTGTCAAAAGATCAGATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTGTT 27339  
QY 301 TGTTCATGTTGATATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTGTT 360  
Db TGTTCATGTTGATATCTCTGTTTGTAGATGCGGCATTTCTGAGGGCTCTGTT 27399  
QY 361 CTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 420  
Db CTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 27459  
QY 421 GGATTGACTTGGCGATGGGCTCTTTTGTGTTTC 455  
Db GGATTGACTTGGCGATGGGCTCTTTTGTGTTTC 27494

RESULT 7  
US-08-966-958-1/c  
; Sequence 1, Application US/08966958  
; Patent No. 5928908  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John  
; APPLICANT: Randes, Matthew  
; TITLE OF INVENTION: METHODS FOR INTRODUCING UNIDIRECTIONAL  
; TITLE OF INVENTION: DELETIONS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brookhaven National Laboratory  
; STREET: P.O. Box 5000  
; CITY: Upton  
; STATE: New York  
; COUNTRY: US  
; ZIP: 11973  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/966,958  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bogosian, Margaret  
; REGISTRATION NUMBER: 25,324  
; REFERENCE/DOCKET NUMBER: AU197-14  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 344-3341  
; TELEFAX: (516) 344-3729  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10754 base pairs  
; TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-966-958-1

Query Match 63.7%; Score 319.2; DB 2; Length 10754;  
Best Local Similarity 94.8%; Pred. No. 3.8e-85;  
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCACATAGGCGTCAGCCAGTTTCTCAGACCCATTATTAATAGGG 213  
DB 10754 CCAGTTTCAGCTTTCATATATAGCGTCAGCCAGTTTCCAGACCCATTATTAATAGGG 10695  
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 273  
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 10635  
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333  
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTACC 10575  
QY 334 AGTACCATGTTGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 393  
DB 10574 AATACCATGCTGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 10515  
QY 394 TGCCCTCAGCTTGTCTTTGGCTAGGATGACTTGGCGATGTGGCTCTTTTGGT 453  
DB 10514 TGCCCTCAGCTTGTCTTTGGCTAGGATGACTTGGCGATGTGGCTCTTTTGGT 10455  
QY 454 TCACATGAACCTTTAAAGTAGTTTTCACAAATTCGTGGAAGAAATGA 501  
DB 10454 TCACATGAACCTTTAAAGTAGTTTTCACAAATTCGTGGAAGAAATGA 10407

RESULT 8  
US-09-215-817-1/c

; Sequence 1, Application US/09215817  
; Patent No. 5968786  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John  
; TITLE OF INVENTION: METHODS FOR INTRODUCING UNIDIRECTIONAL  
; TITLE OF INVENTION: DELETIONS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brookhaven National Laboratory  
; STREET: P.O. Box 5000  
; CITY: Upton  
; STATE: New York  
; COUNTRY: US  
; ZIP: 11973  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/215,817  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/966,958  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bogosian, Margaret  
; REGISTRATION NUMBER: 25,324  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 344-3341  
; TELEFAX: (516) 344-33729  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10754 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-215-817-1

Query Match 63.7%; Score 319.2; DB 2; Length 10754;  
Best Local Similarity 94.8%; Pred. No. 3.8e-85;  
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCACATAGGCGTCAGCCAGTTTCTCAGACCCATTATTAATAGGG 213  
DB 10754 CCAGTTTCAGCTTTCATATATAGCGTCAGCCAGTTTCCAGACCCATTATTAATAGGG 10695  
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 273  
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 10635  
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333  
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTACC 10575  
QY 334 AGTACCATGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 393  
DB 10574 AATACCATGCTGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 10515  
QY 394 TGCCCTCAGCTTGTCTTTGGCTAGGATGACTTGGCGATGTGGCTCTTTTGGT 453  
DB 10514 TGCCCTCAGCTTGTCTTTGGCTAGGATGACTTGGCGATGTGGCTCTTTTGGT 10455  
QY 454 TCACATGAACCTTTAAAGTAGTTTTCACAAATTCGTGGAAGAAATGA 501  
DB 10454 TCACATGAACCTTTAAAGTAGTTTTCACAAATTCGTGGAAGAAATGA 10407

RESULT 9

US-09-342-353-1/c

; Sequence 1, Application US/09342353  
; Patent No. 6248569  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John  
; TITLE OF INVENTION: METHOD FOR INTRODUCING UNIDIRECTIONAL NESTED DELETIONS  
; FILE REFERENCE: CIP OF U.S. Application 08/966,958  
; CURRENT APPLICATION NUMBER: US/09/342,353  
; EARLIER FILING DATE: 1999-06-29  
; EARLIER APPLICATION NUMBER: 08/966,958  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 10754  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-342-353-1

Query Match 63.7%; Score 319.2; DB 4; Length 10754;  
Best Local Similarity 94.8%; Pred. No. 3.8e-85;  
Matches 330; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 154 CCAGTTTCAGCTTTCACATAGGCGTCAGCCAGTTTCTCAGACCCATTATTAATAGGG 213  
DB 10754 CCAGTTTCAGCTTTCATATATAGCGTCAGCCAGTTTCCAGACCCATTATTAATAGGG 10695  
QY 214 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 273  
DB 10694 AATCCTTCCCATCTGTTTCTCAGGTTTGCAAGATCAGATGTTGATAT 10635  
QY 274 GCGGCATTAATTTCTGAGGCGCTGTTCTGTCATTGGTTGATATCTGTTTGTACC 333  
DB 10634 GTGGCGTTAATTTCTGAGGCGCTGTTCTGTCATTGGTTAATCTGTTTGTACC 10575  
QY 334 AGTACCATGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 393  
DB 10407 AATACCATGCTGTTTGTACTGTAAGCTTGTAGTGTAGTTGAAGTACGATGA 10515





OY	3	TTTGATGTTTAAAGCAATGAGTCCCTGTCCTCCATCCATGCTATGCTCGAATGGAATGCTGGG	62
Db	49139	TTTGGATTTTAAATAAATAATCTCTTCCTTAAGCTGATGCCCAACACACATTTTCTAAG	49198
OY	63	TTTTCTCTTAAGGATTTTATGTTTTAGTCTCAACATTTAAGCTTTAAATCCATCTTGAA	122
Db	49199	TTTTCTCTTAAGGATTTCTTATAGTTTCAAAAGCTTATATTTAAGCTTTTAAATCCACCTCAAG	49258
OY	123	TTAATTTTGAATAGGTGAAGGAAGGATCCAGTTAGCTTCTTCAATAGAGGCTAGC	182
Db	49259	TTAATTTTATATATAGTGAAGGAAGCGGGTCTCTGTTCACTCTTTGCAATGGGCCACG	49318
OY	183	CAGTTTTCTCAGCACCATTTATTAATAATAGGAATCCCTTTCCCATTCCTGTTTCTCA	242
Db	49319	CAGCAATCCGAGAACCATTTATGGAATTAAGAAATCTTTCCCATTCGTT - ATTTGTCA	49377
OY	243	GGTTTGCAAAAGATCAGATAGTGTATGATATGGGGCATTAATTTCTAGAGGCTCTGTCTCG	302
Db	49378	ACTTTTGCAAAAGATCGGATAGCTAGTAGAGTGGG - TTTTCTCTGGGTATCTACTCTG	49435
OY	303	TTCCATTTGGTGAATATCTCTGTTTGGTACCAATACATGTTGTTGGTATCTAGAGCC	362
Db	49436	TTACATTTGGCTATAGTGTCTGTTTGTATACATATACATGCGTCTTTTGTACTAATGCGC	49495
OY	363	TTGTAGCTATGTTGAAGTCAAGTACGATGATGCCCTCCAGCTTTGTTCTTTGGCTTAG	422
Db	49496	TCATTAACATAGTTTAAAGTTGGATTAAGTTATGCCG - CTGCTTTGGCTGTTTTTGGCTTAG	49554
OY	423	ATTGACTTTGGCGATGTTGGGCTCTTTTGTGTTGCTCAATGGAATTTTAAAGATTTTTC	482
Db	49555	ATTGCTTTGGCTATGAGGCTCTTTTTCACCTTCATATGATATTTTGAATAGTTTTCCT	49614
OY	483	AATTCCTGTGAAGAAAG	498
Db	49615	AATTCCTTGAAAAATG	49630

```

RESULT 13
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sait, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078.294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contlig
US-09-078-294-3

Query Match          49.6%; Score 248.4; DB 4; Length 80595;
Best Local Similarity 72.8%; Pred. No 9,2e+64;
Matches 361; Conservative 0; Mismatches 131; Indels 4; Gaps 3;

OY      3   TTGCGTGTATAGACATGAAGTCCTTGCCCATCCTATGTCGTAATGGTAGTGCCTGGG 62
        ||| |||||
DB       49401 TTTTGAAATTTAATAAATAATTCTTCCATAGCGTAGCGCCAGACACACTTTCTAGG 49460

OY      63   TTTTCTCTPAGGGATTTATAGTTTTAGGCTCAACATTAAGCTTATACATCTTGAA 122
        ||| |||||
DB       49461 TTTTCTCTAGGATCTTATAGTCAAGCTTATATTTAAGCTTTAATTCACCTCAAG 49520

OY      123  TTAATTTTGTATAGGCTGTAGAAGGAGGATCCAGTTTGAGCTTCTTACATAGGGCTAGC 182
         ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB       49521 TTAATTTTATATAGTGAATAGCAGGGGCTCGTTTCACTCTTTTGGCATGTGGCCAGC 49580
         ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

[illegible]

```

RESULT 14
US-09-146-053-3/c
; Sequence 3, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-3

Query Match          49.3%; Score 246.8; DB 4; Length 50000;
Best Local Similarity 73.4%; Pred. No. 2.3e-63;
Matches 348; Conservative 0; Mismatches 112; Indels 14; Gaps 2;

      1  CTTTTCGTCTTTAGACATGAGTCCTTGCCCATGGCCTATGTCCTGAATGTAATGCCCTG 60
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db       31100 CTTTGGAGACTGATGTAATAATTATTACCTGGGCANAGGCCAAGAATTTCTTA 310414
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

      61  GGTTTCTCTAGGCAATTTATGTTTTAGGCTCAACATTTAAAGTCTTTAATCCATCTTG 120
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db       31040 AGTTTGATCTAGGATTTTATATAGTTTATGTTTGCCTTCAATTTAAGTCTTTATTCACACTTG 309811
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

      121  AATTAATTTTGTATTAAGTGTGAAGCAAGAGGATCCAGTTTCACTTTCATATAGGCTTA 180
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db       30980 TGTTAATTTTGTATGCGGTACAGAGTAAGAGTGTGAATTTCAATCTCTCGCANATAGTTTA 309211
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

      181  GCCAGTTTCTCACACCACCATTTATTAATAGGCAATCT-----TTCCCCATTCCTT 232
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db       30920 GCCAGTTTCCCACACCATTTATGATGAGGGTGTCCTAATGGGTATTCGCCCATTTGTTT 308611
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

      233  GTTTTTCACAGGTTTGCAAGAATGATGATGTTGTAGATATGGGCGATTATTTCTGAGGG 292
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db       30860 ATTTTGTCCACTTGTGCTGAAGATCAAGTTGGTTGTAGGTGATCAAGTTTATTTTCAGGGGT 308011
      ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

      293  CTCGTCTGTCTCCCATTTGGTTGAATCTCTG-----TTTGGATACAGATACCATATGTTGT 346

```

Search completed: July 4, 2003, 19:02:53  
Job time : 37.0233 secs

```
Db      30800 CTCTATTCGTTCCATGTGCTATGTCGTGATTTATTTTGTGACGACACATGCTCT 30741
OY      347  TTTGGTACTGAGCCCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTT 406
Db      30740 TTTGGTACTGAGCCCTGTAGTGTAGTTGAAGTCAGGTAGCATGATGCCCTCCAGCTTT 30681
OY      407  GTCTTTTGGCTTGAAGTATGATGGCATGTGGCTCTTTTGGTTCCACAT 460
Db      30680 GTCTTTTGGCCACAGATGTGCTTGTATCTGTGCGCTTTTGTATTCATAT 30627
```

## RESULT 15

```
US-09-734-675-3/c
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE: DNA
; ORGANISM: Human
US-09-734-675-3
```

Query Match 49.1%; Score 246.2; DB 4; Length 38844;  
Best Local Similarity 73.3%; Pred. No. 3.1e-63;  
Matches 373; Conservative 0; Mismatches 123; Indels 13; Gaps 4;

```
OY      1  CTTTGGTGTGTTAGACATGAAGTCTGCCCATGCTTATGCTGAATGATGCTG 60
Db      11095 CTTTGGAGACTTACTCTAAATCTGTGCCAAGCTATGCTACAAAGATTTCTTA 11036
OY      61  GGTCTTCTCTAGGATTT-----ATGGTTTGTAGTCTAAACATTTAAGCTTTAATCC 114
Db      11035 GGTCTTCTCTAGGCTTTCTGATTAATTTTGTGCTTACATTTAATCTTTACTGC 10976
OY      115  ATCTGAATTAATTTGTATPAAGGTAGAGGAAGGATCCAGTTTCAGCTTCTACATA 174
Db      10975 ATCTGAGTATGTTTGTATTTAGTATGAGTGA---GTCCAATTTCAATTTTCTGCTTC 10920
OY      175  GGGCTAGCATTCTTCAGACACCATTTAATAATAGGAATCCCTTCCCATGCTTGT 234
Db      10919 TGACTAGCTACTATTCACGACCATTTATGAATGAAAGTCTCTCCCATGCTTAT 10860
OY      235  TTTCTCAGTTTGTCAAGATCAGATGATGATGATGCGCATTTCTGAGGCT 294
Db      10859 TTTTGTTAATTTTATGAAATCTGATGCGCTGTGATGTGCGCTTATTTCTGGGCTCT 10800
OY      295  CTGTCGTGTCATGCTTGAATATCTGTTTGTGACAGTACCATGTTGTTGTTA 354
Db      10799 CTATTCGTCTTATGTTTATGTCGTTTGG-ACCAATGCCATTCGTTTGATTA 10741
OY      355  CTGAGCCTTGTAGTGTAGTTGAAGTCAGTACGATGATGCCCTCCAGCTTGTCTTTT 414
Db      10740 CTGAGCCTTGAAGTATGTTGAAGTCAGATAATGAGGCTCTGTCCTTTT 10681
OY      415  GGCCTAGGATGACTGGCGATGAGGCG--TCCTTTTGTTCACATGAACTTAAGT 472
Db      10680 TTATTAGAGTCTTTGGCTATFAGGCTTTTGTGACCATATGAATAATTACAA 10621
OY      473  AGTTTTCCAATTCGTGAAGAAAGTAA 501
Db      10620 AGCTTTTCTAATCTTGTGTGAAATTA 10592
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.778 Seconds

(without alignments)  
6765.003 Million cell updates/sec

Title: US-10-083-853b-2\_COPY\_29421\_29921

Perfect score: 501  
Sequence: 1 ctttgcgtttagacatg .....cattctgtgaagaagaaatga 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database 8  
1: /SID2/gcgdata/geneq/geneq-n-emb1/NA1980.DAT:\*  
2: /SID2/gcgdata/geneq/geneq-n-emb1/NA1981.DAT:\*  
3: /SID2/gcgdata/geneq/geneq-n-emb1/NA1982.DAT:\*  
4: /SID2/gcgdata/geneq/geneq-n-emb1/NA1983.DAT:\*  
5: /SID2/gcgdata/geneq/geneq-n-emb1/NA1984.DAT:\*  
6: /SID2/gcgdata/geneq/geneq-n-emb1/NA1985.DAT:\*  
7: /SID2/gcgdata/geneq/geneq-n-emb1/NA1986.DAT:\*  
8: /SID2/gcgdata/geneq/geneq-n-emb1/NA1987.DAT:\*  
9: /SID2/gcgdata/geneq/geneq-n-emb1/NA1988.DAT:\*  
10: /SID2/gcgdata/geneq/geneq-n-emb1/NA1989.DAT:\*  
11: /SID2/gcgdata/geneq/geneq-n-emb1/NA1990.DAT:\*  
12: /SID2/gcgdata/geneq/geneq-n-emb1/NA1991.DAT:\*  
13: /SID2/gcgdata/geneq/geneq-n-emb1/NA1992.DAT:\*  
14: /SID2/gcgdata/geneq/geneq-n-emb1/NA1993.DAT:\*  
15: /SID2/gcgdata/geneq/geneq-n-emb1/NA1994.DAT:\*  
16: /SID2/gcgdata/geneq/geneq-n-emb1/NA1995.DAT:\*  
17: /SID2/gcgdata/geneq/geneq-n-emb1/NA1996.DAT:\*  
18: /SID2/gcgdata/geneq/geneq-n-emb1/NA1997.DAT:\*  
19: /SID2/gcgdata/geneq/geneq-n-emb1/NA1998.DAT:\*  
20: /SID2/gcgdata/geneq/geneq-n-emb1/NA1999.DAT:\*  
21: /SID2/gcgdata/geneq/geneq-n-emb1/NA2000.DAT:\*  
22: /SID2/gcgdata/geneq/geneq-n-emb1/NA2001A.DAT:\*  
23: /SID2/gcgdata/geneq/geneq-n-emb1/NA2001B.DAT:\*  
24: /SID2/gcgdata/geneq/geneq-n-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	478.6	95.5	5065	21 AAC69111	Human secreted pro
2	478.6	95.5	5065	22 AAC69111	Human secreted pro
3	477.4	95.3	4709	21 AAC69112	Human secreted pro
4	473.8	94.6	1542	23 AAS76725	DNA encoding novel
5	473.8	94.6	1542	23 AAS76725	DNA encoding novel
6	473.8	94.6	1758	23 AAS74901	DNA encoding novel
7	473.8	94.6	1758	23 AAS75196	DNA encoding novel
8	473.8	94.6	1942	23 AAS80525	DNA encoding novel
9	473.8	94.6	2205	23 AAS72639	DNA encoding novel

C 10	473.8	94.6	3052	23 AAS76197	DNA encoding novel
C 11	473.8	94.6	3105	23 AAS78724	DNA encoding novel
C 12	473.8	94.6	3105	23 AAS78792	DNA encoding novel
C 13	473.8	94.6	3105	23 AAS79645	DNA encoding novel
C 14	473.8	94.6	3239	23 AAS76858	DNA encoding novel
C 15	473.8	94.6	4114	23 AAS77062	DNA encoding novel
C 16	473.8	94.6	4289	23 AAS75564	DNA encoding novel
C 17	473.8	94.6	4502	23 AAS78636	DNA encoding novel
C 18	473.8	94.6	5142	23 AAS68835	DNA encoding novel
C 19	473.8	94.6	5654	23 AAS76524	DNA encoding novel
C 20	473.8	94.6	8779	23 AAS71656	DNA encoding novel
C 21	473.8	94.6	13931	23 AAS72644	DNA encoding novel
C 22	472.2	94.3	1827	23 AAS76723	DNA encoding novel
C 23	472.2	94.3	1096	23 AAS76063	DNA encoding novel
C 24	472.2	94.3	1197	22 AAK77110	DNA encoding novel
C 25	472.2	94.3	1200	22 AAK77109	DNA encoding novel
C 26	472.2	94.3	1248	23 AAS69649	DNA encoding novel
C 27	472.2	94.3	1248	23 AAS74285	DNA encoding novel
C 28	472.2	94.3	1248	23 AAS76231	DNA encoding novel
C 29	472.2	94.3	1569	23 AAS68463	DNA encoding novel
C 30	472.2	94.3	1569	23 AAS71273	DNA encoding novel
C 31	472.2	94.3	1821	23 AAS80517	DNA encoding novel
C 32	472.2	94.3	2365	23 AAS73523	DNA encoding novel
C 33	472.2	94.3	2606	23 AAS78958	DNA encoding novel
C 34	472.2	94.3	2940	23 AAS78729	DNA encoding novel
C 35	472.2	94.3	3341	23 AAS68587	DNA encoding novel
C 36	472.2	94.3	3471	23 AAS71209	DNA encoding novel
C 37	472.2	94.3	3579	23 AAS70252	DNA encoding novel
C 38	472.2	94.3	3579	23 AAS73833	DNA encoding novel
C 39	472.2	94.3	4555	23 AAS84104	DNA encoding novel
C 40	472.2	94.3	6334	23 AAS76938	DNA encoding novel
C 41	472.2	94.3	30274	23 AAS85251	DNA encoding novel
C 42	472.2	94.3	167343	24 ABL64403	Stomach cancer rel
C 43	472.2	94.3	167343	24 ABL67239	Thyroid cancer rel
C 44	470.6	93.9	1200	23 AAS76061	DNA encoding novel
C 45	470.6	93.9	1200	23 AAS76691	DNA encoding novel

#### ALIGNMENTS

RESULT 1	
ID AAC69111	standard; DNA; 5065 BP.
AC AAC69111;	
XX	
XX	
XX	31-JAN-2001 (first entry)
XX	
XX	
DE	Human secreted protein gene 3 clone HPJCX13.
KW	Cytostatic; immunosuppressive; nocitropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antidiabetic; antidiabetic;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein; ss.
XX	
OS	Homo sapiens.
XX	
XX	
XX	WO200055371-A1.
PD	21-SEP-2000.
XX	
PF	16-MAR-2000; 2000WO-US06783.
XX	
XX	
XX	18-MAR-1999; 99US-0125055.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
XX	Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI	Komatsu S, G, Lafleur DM, Moore PA, Olsen HS, Young PE;
XX	
XX	WPI; 2000-594448/56.

DR P-PSDB: AAB38011.

XX New nucleic acid molecules encoding 27 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

PS Claim 1: Page 383-384; 453pp: English.

XX  
XX The invention relate to the isolation of genes AAC69084-C69119 encoding  
CC 27 human secreted proteins AAB37984-B38019. The genes can be used to  
CC generate fusion proteins by linking to the gene for the human  
CC immunoglobulin G Fc portion (AAC69075) for increasing the stability of  
CC the fusion protein as compared to the human protein only. The genes and  
CC proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

XX Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other:

Query Match 95.5%; Score 478.6; DB 21; Length 5065;

Best Local Similarity 95.4%; Pred. No. 1.5e-120;

Matches 478; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

OY 1 CTTTGGGTTTACAGAGAGTCCGCCATGCGTATGCGATGATGCGTGC 60  
DB 1090 CTTTGGGTTTACAGAGAGTCCGCCATGCGTATGCGATGATGCGTGC 1149  
OY 61 GGTTCCTCTAGAGGATTTTATGTTTATGCTATACATTTAAGCTTTAATCCATCTG 120  
DB 1150 GGTTCCTCTAGAGGATTTTATGTTTATGCTATACATTTAAGCTTTAATCCATCTG 1209  
OY 121 AATTATTTTGTATAGGCTGATGAGAGGATCCAGTTTACGCTTTACATAGGCTTA 180  
DB 1210 AATTATTTTGTATAGGCTGATGAGAGGATCCAGTTTACGCTTTACATAGGCTTA 1269  
OY 181 GCCAGTTTTCACAGCACCATTATTAATAGGAATCCTTCCCATGCGTGTCTTCT 240  
DB 1270 GCCAGTTTTCACAGCACCATTATTAATAGGAATCCTTCCCATGCGTGTCTTCT 1329  
OY 241 CAGGTTTGTCAAGATCAGATAGTGTATAGGAGGATATTTCTGAGGCTCTCTTC 300  
DB 1330 CAGGTTTGTCAAGATCAGATAGTGTATAGGAGGATATTTCTGAGGCTCTCTTC 1389  
OY 301 TGTTCATTTGTTGATATCTCTGTTTGTGACCACTACCAATGTTTGTGTTACTGTA 360  
DB 1390 TGTTCATTTGTTGATATCTCTGTTTGTGACCACTACCAATGTTTGTGTTACTGTA 1449  
OY 361 CCGTAGTGTATGTTGAGTACAGTACGATCCCTCCAGCTTGTGTTGGCTTA 420  
DB 1450 CCGTAGTGTATGTTGAGTACAGTACGATCCCTCCAGCTTGTGTTGGCTTA 1509  
OY 421 GGATTCAGTTCGCGATGTGGCTCTTTTGTGTTCCACATGAACCTTAAAGTAGTTT 480  
DB 1510 GGATTCAGTTCGCGATGTGGCTCTTTTGTGTTCCACATGAACCTTAAAGTAGTTT 1569  
OY 481 CCAATTCTGTGAGAGAGTAA 501  
DB 1570 CCAATTCTGTGAGAGAGTCA 1590

RESULT 2  
AAB33219

ID AAB33219 standard; CDNA; 5065 BP.

AC AAB33219;

DT 23-MAR-2001 (first entry)

DE Human secreted protein gene 7 SEQ ID NO:17.

KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;  
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;  
KW immunostimulant; cytosolic; cardiac; vascular; anti-angiogenic;  
KW ophthalmological; neuroprotectant; neotropic; anticonvulsant; viderary;  
KW antialzheimers; antiparkinsonian; antimicrobial; immune disorder;  
KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;  
KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;  
KW cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;  
KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;  
KW corneal graft neovascularisation; neurological disorder; regeneration;  
KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;  
KW infectious disease; chemotaxis; ss.

OS Homo sapiens.

PN WO200076530-A1.

PD 21-DEC-2000.

PF 01-JUN-2000; 2000WO-US14933.

PR 11-JUN-1999; 99US-0138572.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsuolis GA;

XX WPI: 2001-071147/08.

DR P-PSDB: AAB64888.

XX Nucleic acids encoding 49 human secreted polypeptides, useful for  
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
PT disease and diabetic retinopathy -

PS Claim 1: Page 453-455; 554pp; English.

CC The polynucleotide sequences given in AAB33213 to AAB33261 encode the  
CC human secreted proteins given in AAB64882 to AAB64930. AAB64931 to  
CC AAB64991 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;  
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;  
CC vascular; antimicrobial; anti-angiogenic; ophthalmological;  
CC neuroprotectant; anticonvulsant; neotropic; antialzheimers;  
CC antiparkinsonian; and vulnerary. The polynucleotides and polypeptides can  
CC be used in the prevention, diagnosis and treatment of diseases associated  
CC with inappropriate polypeptide expression. Disorders that may be  
CC prevented, diagnosed and/or treated by the above methods include immune  
CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and  
CC human immuno-deficiency virus (HIV) infections), hyperproliferative  
CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary  
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
CC neovascularisation and diabetic retinopathy), neurological disorders  
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
CC infectious diseases and/or for promoting wound healing, regeneration and  
CC /or chemotaxis. AAB33204 to AAB33212 and AAB64881 represent sequences  
CC used in the exemplification of the present invention.

XX Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other:

Query Match 95.5%; Score 478.6; DB 22; Length 5065;

Best Local Similarity 95.4%; Pred. No. 1.5e-120;  
Matches 478; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

```
OY 1 CTTTGGGCTTTAGACATGAAGTCCTTCCCATGCCATATCCGATGCTATGCTG 60
    |||||||
DB 1090 CTTTGGGCTTTAGACATGAAGTCCTTCCCATGCCATATCCGATGCTATGCTG 1149
    |||||||
OY 61 GGTTCCTTACAGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
    |||||||
DB 1150 GGTTCCTTACAGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1209
    |||||||
OY 121 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTACGTTTCTACATAGGCTA 180
    |||||||
DB 1210 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTACGTTTCTACATAGGCTA 1269
    |||||||
OY 181 GCCAGTTTTCACAGACATTTATTAATTAAGGAAATCCTTCCCATGCTTGTCTTCT 240
    |||||||
DB 1270 GCCAGTTTTCACAGACATTTATTAATTAAGGAAATCCTTCCCATGCTTGTCTTCT 1329
    |||||||
OY 241 CAGGTTTGTCAAAAGATCAGATAGTTGTATGATGCGGATTTTCTGAGGGCTCTGTC 300
    |||||||
DB 1330 CAGGTTTGTCAAAAGATCAGATAGTTGTATGATGCGGATTTTCTGAGGGCTCTGTC 1389
    |||||||
OY 301 TGTTCATTTGGTTATATCTCTGTTTGTGACACAGTACATGTTTGTGTTACTGTAG 360
    |||||||
DB 1390 TGTTCATTTGGTTATATCTCTGTTTGTGACACAGTACATGTTTGTGTTACTGTAG 1449
    |||||||
OY 361 CCTGTGATGTAGTTTGAAGTCAGTACATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
    |||||||
DB 1450 CCTGTGATGTAGTTTGAAGTCAGTACATGATGCCCTCCAGCTTGTCTTTGGCTTA 1509
    |||||||
OY 421 GGATTTGACTTGGCATGTGGGCTCTTTTGTGTCACATCACTTAAAGAGTTTGT 480
    |||||||
DB 1510 GGATTTGACTTGGCATGTGGGCTCTTTTGTGTCACATCACTTAAAGAGTTTGT 1569
    |||||||
OY 481 CCAATTCGTGTAAGAAAGTAA 501
    |||||||
DB 1570 CCAATTCGTGTAAGAAAGTCA 1590
    |||||||

RESULT 3
AAC69112/c
ID AAC69112 standard; DNA; 4709 BP.
AC AAC69112;
XX
XX 31-JAN-2001 (first entry)
DE Human secreted protein gene 3 clone HNHCT15.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antifibrotic; hepatotropic; antidiabetic; antiinflammatory; antileuk;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardial; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ss.
XX
XX Homo sapiens.
OS
XX WO200055371-A1.
PN
XX 21-SEP-2000.
PD
XX 16-MAR-2000; 2000WO-US06783.
PF
XX 18-MAR-1999; 99US-0125055.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Ni J, Edner R, Rosen CA, Shi Y, Birse C, Florence K;
PI Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX WPI: 2000-594448/56.
DR P-PSDB; AAB38012.
```

XX New nucleic acid molecules encoding 27 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PI used as food additives or preservatives -  
PS  
XX Claim 1; Page 385-386; 453pp; English.

XX The invention relate to the isolation of genes AAC69084-C69119 encoding  
CC 27 human secreted proteins AAB37984-B38019. The genes can be used to  
CC generate fusion proteins by linking to the gene for the human  
CC immunoglobulin G Fc portion (AAC69075) for increasing the stability of  
CC the fusion protein as compared to the human protein only. The genes and  
CC proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

XX Sequence 4709 BP; 1868 A; 990 C; 836 G; 939 T; 76 other;

Query Match 95.3%; Score 477.4; DB 21; Length 4709;

Best Local Similarity 95.4%; Pred. No. 3.2e-120;

Matches 478; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

```
OY 1 CTTTGGGCTTTAGACATGAAGTCCTTCCCATGCCATATCCGATGCTATGCTG 60
    |||||||
DB 3654 CTTTGGGCTTTAGACATGAAGTCCTTCCCATGCCATATCCGATGCTATGCTG 3595
    |||||||
OY 61 GGTTCCTTACAGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 120
    |||||||
DB 3594 GGTTCCTTACAGGATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 3535
    |||||||
OY 121 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTACGTTTCTACATAGGCTA 180
    |||||||
DB 3534 AATTAAATTTTGTATTAAGCTGTAAAGAGGATCCAGTTTACGTTTCTACATAGGCTA 3475
    |||||||
OY 181 GCCAGTTTTCACAGACATTTATTAATTAAGGAAATCCTTCCCATGCTTGTCTTCT 240
    |||||||
DB 3474 GCCAGTTTTCACAGACATTTATTAATTAAGGAAATCCTTCCCATGCTTGTCTTCT 3415
    |||||||
OY 241 CAGGTTTGTCAAAAGATCAGATAGTTGTATGATGCGGATTTTCTGAGGGCTCTGTC 300
    |||||||
DB 3414 CAGGTTTGTCAAAAGATCAGATAGTTGTATGATGCGGATTTTCTGAGGGCTCTGTC 3355
    |||||||
OY 301 TGTTCATTTGGTTATATCTCTGTTTGTGACACAGTACATGTTTGTGTTACTGTAG 360
    |||||||
DB 3354 TGTTCATTTGGTTATATCTCTGTTTGTGACACAGTACATGTTTGTGTTACTGTAG 3295
    |||||||
OY 361 CCTGTGATGTAGTTTGAAGTCAGTACATGATGCCCTCCAGCTTGTCTTTGGCTTA 420
    |||||||
DB 3294 CCTGTGATGTAGTTTGAAGTCAGTACATGATGCCCTCCAGCTTGTCTTTGGCTTA 3235
    |||||||
OY 421 GGATTTGACTTGGCATGTGGGCTCTTTTGTGTCACATCACTTAAAGAGTTTGT 480
    |||||||
DB 3234 GGATTTGACTTGGCATGTGGGCTCTTTTGTGTCACATCACTTAAAGAGTTTGT 3175
    |||||||
OY 481 CCAATTCGTGTAAGAAAGTAA 501
    |||||||
DB 3174 CCAATTCGTGTAAGAAAGTCA 3154
    |||||||

RESULT 4
AAS67689/c
ID AAS67689 standard; cDNA; 1542 BP.
```

XX	AA67689;
AC	
XX	13-FEB-2002 (first entry)
DT	
DE	DNA encoding novel human diagnostic protein #3493.
XX	
XX	Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drimanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
XX	
DR	P-PsDB; ABG03502.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 1; SEQ ID No 3493; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. *Ab564197-AMS94564* represent novel human diagnostic coding sequences of the invention.  
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Query Match	94.68;	Score 473.8;	DB 23;	Length 1542;
Best Local Similarity	96.68;	Pred. No. 2.1e-119;		
Matches 484; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0

QY 1CTTTTGGGTTTATACATGAAGTCGCTGGCCAGCCCTATGTCGTGAATGGATTGCGCG 60  
Db 1432 CTTTGGTGCTTTGAGACATGAAGTCCTTCCCAACCCCTATGTCCTGAATGGATTGCCCA 137  
QY 61 GGTTCCTCTTCAAGGCAATTTATGTTTAAAGTCTCAACATTTAAGCTTTATATCCATCTTG 120  
Db 1372 GGTTCCTCTTCAAGGCTTTTATAGTTTAAAGTCTCAAGCTTTAAGCTTTAATCCATCTTG 131  
QY 121 AATTAAATTTTGTATAAAGGTGAAGGAAGGATCCAGTTTCAGCTTTCTTACATAGGGCTA 180

Db	1312	AATTAATTTTGTATAGGCTGAAGGAGGAGATCCAGTTTCAGCTTCTCATATATGCGTA	1253
QY	181	GCACGTTTTCTCAGACACCATTATTAATAATAGGAAATCCTTTCGCCATGCTGTGTTTTCT	240
Db	1252	GCCAGTTTTTCCACGACCATTATTAATAATAGGAGATCCTTTCGCCATGCTGTGTTTTGT	1193
QY	241	CAGTTTGTCAAGATCAGATAGTAGTTAGATATATGCGGACATTATTTCTGAGGGCTCTTTC	300
Db	1192	CAGTTTGTCAAGATCAGATAGTAGTTAGATATATGCGGACATTATTTCTGAGGGCTCTTTC	1133
QY	301	TGTTCCATTTGGTGAATCTCTCTGTTTTGGTACCAAGTACCATGTGTTTTGGTACTAGTAG	360
Db	1132	TGTTCCATTTGAATCATATCTCTGTTTTGGTACCAAGTACCATGTGTTTTGGTACTAGTAG	1073
QY	361	CCTTTGTAGTGTAGTTGAAGTACAGATAGCATGATGCTCCAGCTTGTCTTTTGGCTTA	420
Db	1072	CCTTTGTAGTGTAGTTGAAGTACAGATAGCATGATGCTCCAGCTTGTCTTTTGGGCTTA	1013
QY	421	GGATGTACTGGGAGATCGGGCTCTTTTGGTCCCATGAACCTTAAGTAGTGTTTT	480
Db	1012	GGATGTACTGGGAGATCGGGCTCTTTTGGTCCCATGAACCTTAAGTAGTGTTTT	953
QY	481	CCAATTCTGTGAAGAAAGTAA	501
Db	952	CCAATTCTGTGAAGAAAGTCA	932

RESULT 5	
AA576725/c	
ID	AA576725 standard; cDNA; 1542 BP.
XX	
AC	AA576725;
XX	
DT	13-FEB-2002 (first entry)

DE	DNA encoding novel human diagnostic protein #12529.
XX	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
XX	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dzmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
XX	
DR	P-PsDB; ABG12538.
XX	

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
PS Claim 1, SEQ ID No 12529, 103pp: English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or



CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 1542 BP; 629 A; 304 C; 303 G; 306 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1542;  
Best Local Similarity 96.6%; Pred. No. 2.1e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTGTTAGACATGAACTCCCTGGCCCATGCTATGCTGATGATGCTG 60
    |||||||
DB 1432 CTTTGGTGTGTTAGACATGAACTCCCTGGCCCATGCTATGCTGATGATGCTG 1373
OY 61 GGTTCCTCTAGAGGATTTTATGTTTATAGCTTAACATTTAAGCTTTAATCCATCTG 120
    |||||||
DB 1372 GGTTCCTCTAGAGGATTTTATGTTTATAGCTTAACATTTAAGCTTTAATCCATCTG 1313
OY 121 AATTAAATTTTGTATAGGTGTAAGGAGGAGGATCCAGTTTCACGTTTCTACATAGGCTTA 180
    |||||||
DB 1312 AATTAAATTTTGTATAGGTGTAAGGAGGAGGATCCAGTTTCACGTTTCTACATAGGCTTA 1253
OY 181 GCCAGTTTCTCGACACATTTTAAATAGGAACCTTCCCATATGCTTTTCTTCT 240
    |||||||
DB 1252 GCCAGTTTCTCGACACATTTTAAATAGGAACCTTCCCATATGCTTTTCTTCTTGT 1193
OY 241 CAGTTTGTCAAAAGATCAGATAGCTTATAGATAGCGCATTTATTCGAGGGCTCTGTTTC 300
    |||||||
DB 1192 CAGTTTGTCAAAAGATCAGATAGCTTATAGATAGCGCATTTATTCGAGGGCTCTGTTTC 1133
OY 301 TGTTCATATGTTGATATCTCTGTTTGGTACAGTACATGCTGTTTGGTACGTTAG 360
    |||||||
DB 1132 TGTTCATATGATATATCTCTGTTTGGTACAGTACATGCTGTTTGGTACGTTAG 1073
OY 361 CCTGTAGTGTAGTGTGAAGTACAGTACATGATGCTGCTCAGCTTTGTTGGCTTA 420
    |||||||
DB 1072 CCTGTAGTGTAGTGTGAAGTACAGTACATGATGCTGCTCAGCTTTGTTGGCTTA 1013
OY 421 GGATGACTTGGGATGTGGCTCTTTTGTGTTCCACATGAACTTTAAAGTAGTTT 480
    |||||||
DB 1012 GGATGACTTGGGATGTGGGCTCTTTTGTGTTCCATATGAACTTTAAAGTAGTTT 953
OY 481 CCAATCTCTGGAAGAAAGTAA 501
    |||||||
DB 952 CCAATCTCTGGAAGAAAGTAA 932
```

## RESULT 6

AAS74901/C  
ID AAS74901 standard; cDNA; 1758 BP.

AC AAS74901;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #10705.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KX food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN W0200175067-A2.

```
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG10714.
```

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID NO 10705; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 1758 BP; 676 A; 381 C; 332 G; 369 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 1758;  
Best Local Similarity 96.6%; Pred. No. 2.2e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTGTTAGACATGAACTCCCTGGCCCATGCTATGCTGATGATGCTG 60
    |||||||
DB 1159 CTTTGGTGTGTTAGACATGAACTCCCTGGCCCATGCTATGCTGATGATGCTG 1100
OY 61 GGTTCCTCTAGAGGATTTTATGTTTATAGCTTAACATTTAAGCTTTAATCCATCTG 120
    |||||||
DB 1099 GGTTCCTCTAGAGGATTTTATGTTTATAGCTTAACATTTAAGCTTTAATCCATCTG 1040
OY 121 AATTAAATTTTGTATAGGTGTAAGGAGGAGGATCCAGTTTCACGTTTCTACATAGGCTTA 180
    |||||||
DB 1039 AATTAAATTTTGTATAGGTGTAAGGAGGAGGATCCAGTTTCACGTTTCTACATAGGCTTA 980
OY 181 GCCAGTTTCTCGACACATTTTAAATAGGAACCTTCCCATATGCTTTTCTTCTT 240
    |||||||
DB 979 GCCAGTTTCTCGACACATTTTAAATAGGAACCTTCCCATATGCTTTTCTTCTTCTT 920
OY 241 CAGTTTGTCAAAAGATCAGATAGCTTATAGATAGCGCATTTATTCGAGGGCTCTGTTTC 300
    |||||||
DB 919 CAGTTTGTCAAAAGATCAGATAGCTTATAGATAGCGCATTTATTCGAGGGCTCTGTTTC 860
OY 301 TGTTCATATGTTGATATCTCTGTTTGGTACAGTACATGCTGTTTGGTACGTTAG 360
    |||||||
DB 859 TGTTCATATGTTGATATCTCTGTTTGGTACAGTACATGCTGTTTGGTACGTTAG 800
```

OY		361	CCCTTGTAGTGTAAGTTGAAGTCAGGATACGATGCCCGCAGCTTGGCTTAA	420
Dd		799	CCTGTAGATATAGTTTGAAGTCAGGATACGATGCCCGCAGCTTGGCTTAA	740
OY		421	GGAATTGACTGGCGGATGTGGGCTCTTTTTGGTCCACATGAACTTTAAGTAGT	480
Dd		739	GGATTCACCTGGGCGGATGTGGGCTCTTTTTGGTCCACATGAACTTTAAGTAGT	680
OY		481	CCAATTCTGTGGAAGAAGATA 501	
Dd		679	CCAATTCTATGAGAAGATCA 659	
<b>RESULT 7</b>				
XX	AS75196/c			
ID	AAST5196 standard; cDNA; 1758 BP.			
XX	AAST5196;			
DT	13-FEB-2002 (first entry)			
DE	DNA encoding novel human diagnostic protein #11000.			
KX	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss,			
XX	Homo sapiens.			
OS				
PX	NO200175067-A2.			
PN				
PD	11-OCT-2001.			
PF	30-MAR-2001; 2001WO-US08631.			
PR	31-MAR-2000; 2000US-0540217.			
PR	23-AUG-2000; 2000US-0649167.			
XX	(HYSE-) HYSEQ INC.			
PA				
XX				
XX				
PI	Dmanac RT, Liu C, Tang YT;			
DR	WPI; 2001-639362/73.			
PT	P-PADB; ABG11009.			
PT	New isolated polynucleotide and encoded polypeptides, useful in			
PT	diagnostics, forensics, gene mapping, identification of mutations			
PT	responsible for genetic disorders or other traits and to assess			
PT	biodiversity -			
PS	Claim 1; SEQ ID NO 11000; 103bp; English.			
XX	The invention relates to isolated polynucleotide (I) and			
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,			
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome			
CC	and gene mapping, and in recombinant production of (II). The			
CC	polynucleotides are also used in diagnostics as expressed sequence tags			
CC	for identifying expressed genes. (I) is useful in gene therapy techniques			
CC	to restore normal activity of (II) or to treat disease states involving			
CC	(II). (II) is useful for generating antibodies against it, detecting or			
CC	quantitating a polypeptide in tissue, as molecular weight markers and as			
CC	a food supplement. (II) and its binding partners are useful in medical			
CC	imaging of sites expressing (II). (I) and (II) are useful for treating			
CC	disorders involving aberrant protein expression or biological activity.			
CC	The polypeptide and polynucleotide sequences have applications in			
CC	diagnostics, forensics, gene mapping, identification of mutations			
CC	responsible for genetic disorders or other traits to assess biodiversity			
CC	and to produce other types of data and products dependent on DNA and			
CC	amino acid sequences. AAS64197-AAS94564 represent novel human			
CC	diagnostic coding sequences of the invention.			
CC	Note: The sequence data for this patent did not appear in the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			

xx	Sequence	1758 BP; 676 A; 381 C; 332 G; 369 T; 0 other;
sq		
Query Match	94.6%; Score 473.8; DB 23; Length 1758;	
Best Local Similarity	96.6%; Fred. No. 2.2e-119;	
Matches 484; Conservative	0; Mismatches 17; Indels 0; Gaps 0;	
OY	1 CTTTGGGCTTTTGACATGAAGTCCTTGGCCCATGCCATGTGCCTGAATGGTAATGGCCTG	60
Dd	1159 CTTTGGGCTTTTGACATGAAGTCCTTGGCCCATGCCATGTGCCTGAATGGTAATGGCCTA	1100
OY	61 GGTTTTCTCTTAGCGGATTTTATGGTTTTAGGCTTACATTAAAGTCTTAAATCCACTTG	120
Dd	1099 GGTTTTCTCTTAGCGGATTTTATGGTTTTAGGCTTACATTAAAGTCTTAAATCCACTTG	1040
OY	121 AATTAAATTTTGTATAAGGCTTAAGAAGGATCCAGTTTCAGCTTCTACATAGGCTA	180
Dd	1039 AATTAAATTTTGTATAAGGCTTAAGAAGGATCCAGTTTCAGCTTCTACATAGGCTA	980
OY	181 GCCAGTTTTCTCAGCACCACTTATTAATAAGGAATCCTTCCCATTGCTGTGTTTTCT	240
Dd	979 GCCAGTTTTCTCAGCACCACTTATTAATAAGGAATCCTTCCCATTGCTGTGTTTTCT	920
OY	241 CAGATTGTGCAAAATGATGATAGTGTATGATATGCGCATATTTGAGGGCTCTGTC	300
Dd	919 CAGATTGTGCAAAATGATGATAGTGTATGATAGGATATTAATTTGAGGGCTCTGTC	860
OY	301 TGTTCCATTTGGTTGATATCTCTGTTTTGGTACCAGTACATGTTGTTGGTTACTGTAG	360
Dd	859 TGTTCCATTTGGTTGATATCTCTGTTTTGGTACCAGTACATGTTGTTGGTTACTGTAG	800
OY	361 CCTGTAGTGTAGTTTGAAGTCAGGTACATGATGATGCCCTCCAGCTTGTCTTTGGCTTA	420
Dd	799 CCTGTAGTGTAGTTTGAAGTCAGGTACGATGATGCCCTCCAGCTTGTCTTTGGCTTA	740
OY	421 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAACTTTT	480
Dd	739 GGATTGACTTGGCATGTGGGCTCTTTTGGTTCACATGAACCTTAAAGTAACTTTT	680
OY	481 CCAATTCGTGAGAAAGTAA 501	
Dd	679 CCAATTCGTGAGAAAGTCA 659	
<hr/>		
RESULT 8		
AAS80525/C		
ID	AAS80525 standard; cDNA; 1942 BP.	
XX		
XX	AAS80525;	
XX		
DT	13-FEB-2002 (first entry)	
DE		
XX	DNA encoding novel human diagnostic protein #16329.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
OS	Homo sapiens.	
WO	200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PE	30-MAR-2001; 2001WO-US08631.	
PR	31-MAR-2000; 2000US-0540217.	
XX	23-AUG-2000; 2000US-0649167.	
PA	(HYSE-) HYSEQ INC.	
PI	Drmnac RT, Liu C, Tang YT;	
WP	1; 2001-639362/73.	

DR P-PSDB: ABG16338.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID No 16329; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II), or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1942 BP; 754 A; 419 C; 368 G; 401 T; 0 other;  
  
Query Match 94.6%; Score 473.8; DB 23; Length 1942;  
Best Local Similarity 96.6%; Pred. No. 2.3e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
  
OY 1 CTTTGGTGTATAGACATGAAGTCCCTGGCCCATGCTATGCTGAATGGTATGCGTG 60  
DB 1343 CTTTGGTGTATAGACATGAAGTCCCTGGCCCATGCTATGCTGAATGGTATGCGTG 1284  
OY 61 GGTTCCTCTAGAGATTTTANGGTTTAAAGCTCAACATTTAAGCTTTAATCATCTTG 120  
DB 1283 GGTTCCTCTAGAGATTTTANGGTTTAAAGCTCAACATTTAAGCTTTAATCATCTTG 1224  
OY 121 AATTAATTTTGTATAGGTATAGAGATGATCCAGTTTCACCTTCTCATAGAGGCTA 180  
DB 1223 AATTAATTTTGTATAGGTATAGAGATGATCCAGTTTCACCTTCTCATAGAGGCTA 1164  
OY 181 GCCAGTTTCTCAGCACCATTTATTAATAGGAAATCCTTCCCATTTGCTTTTCT 240  
DB 1163 GCCAGTTTCTCAGCACCATTTATTAATAGGAAATCCTTCCCATTTGCTTTTCT 1104  
OY 241 CAGGTTTGTCAAGATCAGATAGTGTATATGCGCATTTATTTCTGAGGCTGCTTC 300  
DB 1103 CAGGTTTGTCAAGATCAGATAGTGTATATGAGCATTTATTTCTGAGGCTGCTTC 1044  
OY 301 TGTTCATGTTGATATCTGTTTGTACACAGATCATGTTGTTTGGTACTGTAG 360  
DB 1043 TGTTCATGTTGATATCTGTTTGTACACAGATCATGTTGTTTGGTACTGTAG 984  
OY 361 CTTTGTAGTGTATGATGAGTCAAGTATGATGCTCCAGCTTTGCTTTTGCGCTTA 420  
DB 983 CTTTGTAGTGTATGATGAGTCAAGTATGATGCTCCAGCTTTGCTTTTGCGCTTA 924  
OY 421 GGATGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTAAGTAGTTT 480  
DB 923 GGATGACTTGGCGATGCGGCTCTTTTGGTCCACATGAACCTTAAGTAGTTT 864  
OY 481 CCAATTCGTGGAAGAAGTAA 501  
DB 863 CCAATTCGTGGAAGAAGTAA 843

RESULT 9  
AAS72639/c  
ID AAS72639 standard; cDNA; 2205 BP.  
XX  
XX AAS72639;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX DNA encoding novel human diagnostic protein #8443.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX P-PSDB; ABC08452.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID No 8443; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II), or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 2205 BP; 928 A; 441 C; 396 G; 440 T; 0 other;  
  
Query Match 94.6%; Score 473.8; DB 23; Length 2205;  
Best Local Similarity 96.6%; Pred. No. 2.4e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
  
OY 1 CTTTGGTGTATAGACATGAAGTCCCTGGCCCATGCTATGCTGAATGGTATGCGTG 60  
DB 1378 CTTTGGAGTTTAAACATGAAGTCCCTGGCCCATGCTATGCTGAATGGTATGCGTG 1319  
OY 61 GGTTCCTCTAGAGATTTTANGGTTTAAAGCTCAACATTTAAGCTTTAATCATCTTG 120  
DB 1378 CTTTGGAGTTTAAACATGAAGTCCCTGGCCCATGCTATGCTGAATGGTATGCGTG 1319

```
Db 1318 GGTTCCTCTTACGGGTTTATAGCTTAAGCTTAACTTAAATCATCTTG 1259
Oy 121 AATTAATTTTGTATAGGTGTAGAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTTA 180
Cc 1258 AATTAATTTTGTATAGGTGTAGAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTTA 1199
Oy 181 GCCAGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 240
Db 1198 GCCAGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 1139
Oy 241 CAGGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 300
Db 1138 CAGGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 1079
Oy 301 TGTTCATTTGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 360
Db 1078 TGTTCATTTGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 1019
Oy 361 CCTTGATGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 420
Db 1018 CCTTGATGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 959
Oy 421 GGATTTGACTTGGCGATGCGCTCTTTTGGTCCACATGAATTTAAAGATGTTTCT 480
Db 958 GGATTTGACTTGGCGATGCGCGCTCTTTTGGTCCACATGAATTTAAAGATGTTTCT 899
Oy 481 CCAATTCGTGTGAAGAAAGTAA 501
Db 898 CCAATTCGTGTGAAGAAAGTAA 878

RESULT 10
AAS76197/c
ID AAS76197 standard; cDNA; 3052 BP.
AC AAS76197;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #12001.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG12010.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1: SEQ ID No 12001; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
```

```
Cc polynucleotides are also used in diagnostics as expressed sequence tags
Cc for identifying expressed genes. (I) is useful in gene therapy techniques
Cc to restore normal activity of (II) or to treat disease states involving
Cc (II). (II) is useful for generating antibodies against it, detecting or
Cc quantitating a polypeptide in tissue, as molecular weight markers and as
Cc a food supplement. (II) and its binding partners are useful in medical
Cc imaging of sites expressing (II). (I) and (II) are useful for treating
Cc disorders involving aberrant protein expression or biological activity.
Cc The polypeptide and polynucleotide sequences have applications in
Cc diagnostics, forensics, gene mapping, identification of mutations
Cc responsible for genetic disorders or other traits to assess biodiversity
Cc and to produce other types of data and products dependent on DNA and
Cc amino acid sequences. AAS76197-AAS94564 represent novel human
Cc diagnostic coding sequences of the invention.
Cc Note: The sequence data for this patent did not appear in the printed
Cc specification, but was obtained in electronic format directly from WIPO
Cc at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3052 BP; 1044 A; 704 C; 639 G; 665 T; 0 other;
XX
Query Match 94.6%; Score 473.8; DB 23; Length 3052;
Best Local Similarity 96.6%; Pred. No. 2,7e-119;
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Oy 1 CTTTGGCTTTTAAACATGAAGCTCCCATCCATGCTTCCGTAATGATTCCTG 60
Db 1717 CTTTGGCTTTTAAACATGAAGCTCCCATCCATGCTTCCGTAATGATTCCTG 1658
Oy 61 GGTTTCTCTAGGAGATTTATGTTTATAGCTTACATTTTAAAGCTTTATCATCTTG 120
Db 1657 GGTTTCTCTAGGAGATTTTATGTTTATAGCTTACATTTTAAAGCTTTATCATCTTG 1598
Oy 121 AATTAATTTTGTATAGGTGTAGAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTTA 180
Db 1597 AATTAATTTTGTATAGGTGTAGAGGAAGGATCCAGTTTCAGCTTCTACATAGGCTTA 1538
Oy 181 GCCAGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 240
Db 1537 GCCAGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 1478
Oy 241 CAGGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 300
Db 1477 CAGGTTTTCACAGACCATTTATTAATAGGAAATCCCTTCCCATTCCTGTTTCT 1418
Oy 301 TGTTCATTTGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 360
Db 1417 TGTTCATTTGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 1358
Oy 361 CCTTGATGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 420
Db 1357 CCTTGATGTTGATATCTCTGTTTGTGTACACAGTACATGTTTGTGTACGTAG 1298
Oy 421 GGATTTGACTTGGCGATGCGCTCTTTTGGTCCACATGAATTTAAAGATGTTTCT 480
Db 1297 GGATTTGACTTGGCGATGCGCTCTTTTGGTCCACATGAATTTAAAGATGTTTCT 1238
Oy 481 CCAATTCGTGTGAAGAAAGTAA 501
Db 1237 CCAATTCGTGTGAAGAAAGTAA 1217

RESULT 11
AAS78724/c
ID AAS78724 standard; cDNA; 3105 BP.
AC AAS78724;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #14528.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
```



CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SO Sequence 3105 BP; 1067 A; 721 C; 658 G; 659 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3105;  
Best Local Similarity 96.6%; Pred. No. 2.7e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCTGATGCTGATGCTG 60
D 1717 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCTGATGCTGATGCTG 1658
QY 61 GGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 120
D 1657 GGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 1598
QY 121 AATTAAATTTTGTATAGGTTGTAAGAGAGGATCCAGTTTCAGCTTCTACATAGGCTA 180
D 1597 AATTAAATTTTGTATAGGTTGTAAGAGAGGATCCAGTTTCAGCTTCTACATAGGCTA 1538
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATGCTTCCCATTTGCTTCT 240
D 1537 GCCAGTTTCTCAGCACCATTATTAATAGGGAATGCTTCCCATTTGCTTCT 1478
QY 241 CAGGTTTGTCAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGTC 300
D 1477 CAGGTTTGTCAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGTC 1418
QY 301 TGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 360
D 1417 TGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 1358
QY 361 CCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 420
D 1357 CCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1298
QY 421 GGATTTGACTTGGCAGATGCTGCTTTTGTGTTCCATGATGATTTTAAAGTACTGTTT 480
D 1297 GGATTTGACTTGGCAGATGCTGCTTTTGTGTTCCATGATGATTTTAAAGTACTGTTT 1238
QY 481 CCAATTCGTGAGAGAGTAA 501
D 1237 CCAATTCGTGAGAGAGTAA 1217
```

RESULT 13

AA579645/C  
ID AA579645 standard; cDNA: 3105 BP.

AA579645;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #15449.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN MO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
DR P-PSDB; ABG15458.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 15449; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pcl\_sequences.

SO Sequence 3105 BP; 1067 A; 721 C; 658 G; 659 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3105;  
Best Local Similarity 96.6%; Pred. No. 2.7e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCTGATGCTGATGCTG 60
D 1717 CTTTGGGTTTGTAGACATGAGTCCTGGCCATGCTGATGCTGATGCTG 1658
QY 61 GGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 120
D 1657 GGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 1598
QY 121 AATTAAATTTTGTATAGGTTGTAAGAGAGGATCCAGTTTTCAGAGGCTCTGTC 180
D 1597 AATTAAATTTTGTATAGGTTGTAAGAGAGGATCCAGTTTTCAGAGGCTCTGTC 1538
QY 181 GCCAGTTTCTCAGCACCATTATTAATAGGGAATGCTTCCCATTTGCTTCT 240
D 1537 GCCAGTTTCTCAGCACCATTATTAATAGGGAATGCTTCCCATTTGCTTCT 1478
QY 241 CAGGTTTGTCAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGTC 300
D 1477 CAGGTTTGTCAAGATCAGATAGTGTAGATATCGGCAATTTTTCAGAGGCTCTGTC 1418
QY 301 TGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 360
D 1417 TGTTCCTCTAGGAGATTTATAGGTTTACGCTACATTTTATCCATCTG 1358
QY 361 CCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 420
D 1357 CCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1298
QY 421 GGATTTGACTTGGCAGATGCTGCTTTTGTGTTCCATGATGATTTTAAAGTACTGTTT 480
D 1297 GGATTTGACTTGGCAGATGCTGCTTTTGTGTTCCATGATGATTTTAAAGTACTGTTT 1238
```



QY 481 CCAATTCGTGAGAAAGTAA 501  
|||||  
Db 1237 CCAATTCGTGAGAAAGTCA 1217

RESULT 14  
AAS76858/c  
ID AAS76858 standard; cDNA; 3239 BP.  
XX  
AC AAS76858;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #12662.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABG12671.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID NO 12662; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3239 BP; 1091 A; 779 C; 686 G; 683 T; 0 other;

Query Match 94.6%; Score 473.8; DB 23; Length 3239;  
Best Local Similarity 96.6%; Pred. No. 2.7e-119;  
Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CTTTGGTGTGTTAGACATGAAGTCCTTCCCATGCTATGTCCTGAATGATATGCCGTG 60  
|||||

Db 1717 CTTTGGTGTGTTGACATGAAGTCCTTCCCATGCTATGTCCTGAATGATATGCCTA 1658  
QY 61 GGTTCCTCTAGGAGTTTATAGTTCATCAATTTAAGTCTTAATCCATCTTG 120  
|||||  
Db 1657 GGTTCCTCTAGGAGTTTATAGTTCATCAATTTAAGTCTTAATCCATCTTG 1598  
|||||  
QY 121 AATTAAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGCGTA 180  
|||||  
Db 1597 AATTAAATTTTGTATAGGTGTAAGGAAGGATCCAGTTTCAGCTTTTACATAGGCGTA 1538  
|||||  
QY 181 GCCAGTTTTCAGCAGCATTATTAATAGGAATCCCTTCCCATGCTTGTTCCT 240  
|||||  
Db 1537 GCCAGTTTTCAGCAGCATTATTAATAGGAATCCCTTCCCATGCTTGTTCCT 1478  
|||||  
QY 241 CAGGTTTGCAGAAAGATACATAGTGTATATATGCGGATTAATTTGTAGGCGCTGTTTC 300  
|||||  
Db 1477 CAGGTTTGCAGAAAGATACATAGTGTATATATGCGGATTAATTTGTAGGCGCTGTTTC 1418  
|||||  
QY 301 TGTTCATGTTGGTATATCTGTTTGGTACCAGTACATGTTGTTGGTACGTAG 360  
|||||  
Db 1417 TGTTCATGTTGGTATATCTGTTTGGTACCAGTACATGTTGTTGGTACGTAG 1358  
|||||  
QY 361 CCTTGATGTTAGTTGAAGTCAGTAGCATGATGCCCTCAGCTTTGTTGGCTTA 420  
|||||  
Db 1357 CCTTGATGTTAGTTGAAGTCAGTAGCATGATGCCCTCAGCTTTGTTGGCTTA 1298  
|||||  
QY 421 GGATTCACCTTGGCATGTCGCTTTTGGTCCACATGAACCTTAAGATGTTTTT 480  
|||||  
Db 1297 AGATTGACTTGGCAAGTGGGCTTTTGGTCCATATGAACCTTAAGATGTTTTT 1238  
|||||

QY 481 CCAATTCGTGAGAAAGTAA 501  
|||||  
Db 1237 CCAATTCGTGAGAAAGTCA 1217

RESULT 15  
AAS77062/c  
ID AAS77062 standard; cDNA; 4114 BP.  
XX  
AC AAS77062;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #12866.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABG12875.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID NO 12866; 103pp; English.  
XX



CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 4114 BP: 1438 A; 951 C; 808 G; 917 T; 0 other;

Query Match: 94.6%; Score 473.8; DB 23; Length 4114;

Best Local Similarity 96.6%; Pred. No. 2.9e-119;

Matches 484; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
OY 1 CTTTGGTGTATTAAGATGAGTCCCTGCGCATATGCTGATGATGATGCGTG 60
DB 2587 CTTTGGTGTATTAAGATGAGTCCCTGCGCATATGCTGATGATGCGTG 2528
OY 61 GGTTCCTCAGAGGATTTTAAAGTCTTAACTTAACTTAACTTAACTTAA 120
DB 2527 GGTTCCTCAGAGGATTTTAAAGTCTTAACTTAACTTAACTTAACTTAA 2468
OY 121 AATTAAATTTTGTATTAAGTGTAAAGAGGATCCAGTTTCAGTTTCTACATAGGCTA 180
DB 2467 AATTAAATTTTGTATTAAGTGTAAAGAGGATCCAGTTTCAGTTTCTACATAGGCTA 2408
OY 181 GCCAGTTTTCACAGCACCATTATTAATAAGGAAATCCTTCCCATGCTGTTTCT 240
DB 2407 GCCAGTTTTCACAGCACCATTATTAATAAGGAAATCCTTCCCATGCTGTTTCT 2348
OY 241 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGCGGCAATATTTCTGAGGGCTCTGTT 300
DB 2347 CAGTTTGTCAAAAGATCAGATAGTGTAGATATGCGGCAATATTTCTGAGGGCTCTGTT 2288
OY 301 TGTTCATTTGTTGATATCTGTTTGTGTACCATGATGTTGTTGTTACTGTAG 360
DB 2287 TGTTCATTTGTTGATATCTGTTTGTGTACCATGATGTTGTTGTTACTGTAG 2228
OY 361 CCTGTAGTGTAGTTGAAGTACAGTACATGATGCTCCAGCTTGTGTTTGGCTTA 420
DB 2227 CCTGTAGTGTAGTTGAAGTACAGTACATGATGCTCCAGCTTGTGTTTGGCTTA 2168
OY 421 GGATTGACTTGCGGATGTTGTTTGTGTTTCCACATGAACTTTAAAGTACTTTT 480
DB 2167 AGATTGACTTGCGGATGTTGTTTGTGTTTCCACATGAACTTTAAAGTACTTTT 2108
OY 481 CCAATTCGTGAGAAAGTAA 501
DB 2107 CCAATTCGTGAGAAAGTCA 2087
```

Search completed: July 4, 2003, 15:25:06  
Job time : 168.778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:09:32 ; Search time 1095.73 Seconds

(without alignments)  
7405.060 Million cell updates/sec

Title: US-10-083-853b-2\_COPY\_18900\_19400

Perfect score: 501  
Sequence: Tttctatgactaaagacac.....gcctccctgttctctttt 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:\*

1: em\_estda:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	9.4	1101	17	CNS008WC
2	41.6	8.3	1101	17	AL052719 Drosophila
3	40	8.0	1101	17	AL098379 Drosophila
4	39.4	7.9	1101	17	BB646396 Drosophila
5	39.2	7.8	274	9	AV223573 Drosophila
6	38.8	7.7	497	12	BF423025 Drosophila

C 7	38.8	7.7	574	17	BH539949
C 8	38.8	7.7	885	13	B1546960
C 9	38.8	7.7	993	17	CNS0256T
C 10	38.6	7.7	414	13	B3936625
C 11	38.6	7.7	641	13	B1585460
C 12	38.6	7.7	836	17	A2682240
C 13	38.6	7.7	883	17	A2690387
C 14	38.6	7.7	921	17	BH150526
C 15	38.2	7.6	425	17	BH183755
C 16	38.2	7.6	840	17	A0027476
C 17	38	7.6	837	17	BH165379
C 18	38	7.6	1110	17	CNS01EXV
C 19	37.8	7.5	709	12	B6385089
C 20	37.6	7.5	288	10	BH181383
C 21	37.6	7.5	1204	17	CNS016E2
C 22	37.4	7.5	493	13	B1946460
C 23	37.4	7.5	559	13	B1574743
C 24	37.4	7.5	624	17	A2415638
C 25	37.4	7.5	698	13	B1584493
C 26	37.4	7.5	1024	17	CNS005BT
C 27	37.2	7.4	600	13	B1988264
C 28	37.2	7.4	898	17	CNS01POH
C 29	37	7.4	636	13	BH184301
C 30	37	7.4	647	17	AG065239
C 31	36.8	7.3	414	17	A0045487
C 32	36.8	7.3	491	10	BH781797
C 33	36.8	7.3	828	17	BH689601
C 34	36.8	7.3	919	17	A2538973
C 35	36.6	7.3	424	17	BH352158
C 36	36.6	7.3	480	17	BH763634
C 37	36.6	7.3	939	17	BH137105
C 38	36.6	7.3	1201	17	CNS002OT
C 39	36.4	7.3	252	10	BH076733
C 40	36.4	7.3	391	17	A2410995
C 41	36.4	7.3	500	9	AU088252
C 42	36.4	7.3	695	17	BH161708
C 43	36.4	7.3	870	17	A2677313
C 44	36.4	7.3	919	17	CNS005RL
C 45	36.4	7.3	996	17	CNS00FHL

## ALIGNMENTS

RESULT 1	CNS008WC	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	BACR18L14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
DEFINITION	AL052719.1 GI:4934268				
ACCESSION	AL052719				
VERSION	GSS.				
KEYWORDS	Drosophila melanogaster.				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)				

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

Fax: 01-555-900-2200  
 Email: genome-res@qsc.riken.go.jp,  
 URL: <http://genome.qsc.riken.go.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wathiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamamaki, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

mouse tissues.

**FEATURES**  
**source**

BASE COUNT  
ORIGIN

189 a      84 c      132 g      154 t

```
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db_xref="taxon:10090"  
/clone="B930009j12"  
/clone_1lb="RIKEN full-length enriched, 10 days neonate  
cerebellum"  
/sex="mixed"  
/tissue_type="cerebellum"  
/dev_stage="10 days neonate"  
/lab_host="DH10B"  
/note="Site_1: SalI; Site_2: BamHI; CDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand CDNA was  
primed with a primer [5'  
GAGGAGAGAGAGATCCACGACTCTTTTGTGGTTTTTNN 3']  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. CDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand CDNA was prepared with the primer adapter of  
sequence [5' GAGGAGAGAGATTTCGAGTTAATTAATATCCCCCCCCCCC  
3']. CDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:  
BamHI"
```

Query Match	8.0%:	Score 40:	DB 10:	Length 559:
Best Local Similarity	52.4%:	Pred. No. 1.3:		
Matches 86:	Conservative	0:	Mismatches 80:	Indels 0:
QY	14	AAGCACCACATATTTTCAAGCTAATCATTAATTTACTTGGAATTCATTCGCAATTA	73	
Db	394	AAGAACTCAGATTAATCTTTAAGGGGCTGGCCCATTAATTTCTCCATGTCAGAGAAATAT	335	
QY	74	TTGTGTTACATTAATTTAGATTTGATTTTCGTTGTTCCAAAGTCATTCATTAATTCATCA	133	
Db	334	ATGGGCAACATTAATTTGACCTTGGCTGTAGTCTTCTCCTTTCTACACTTCTCTCTACTA	275	
QY	134	TTTTCGCAGCAAAATCCTTATGTCGTCATACATACATAGATATCTTC	181	
Db	274	TTTTTCTCCACGCTTCTGCTCTCTGCTCTCCATAGCTCATCAATCGTCT	227	

[illegible]

COMMENT

collaborating with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see: [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source

BASE COUNT	291	a	178	c	34	g	216	t	382	others
ORIGIN										
Query Match	7.9%	Score	39.4;	DB	17;	Length	1101;			
Best Local Similarity	26.8%	Pred. NO.	2.4;							
Matches	97;	Conservative	88;	Mismatches	177;	Indels	0;	Gaps	0;	

OY	12	AAAGCACCACATATTTTCCAAAGCTAAGTCATCAATAATTACTTGAATTCATTTGGCAT	71
Dd	554	AAAATMAAMTATAGAMTKTATPAMAATATAMVTAADGTRATKTTTAAAAAAAATTTATWM	613
OY	72	TATTGTGTCATCAATAAATTTAGATTGATTTGCGTTCACAGTCCATCCAATTWCCATC	131
Dd	614	TTAAATMTAGTGIDATPTAMATTTATRWMTAMATTTATTTVAAAADTTGTGTTAAAAANW	673
OY	132	CATTGGCAGCCAAATCCTCTTATTTGGTCATACATACAGTAGATCTTCCACAGAAC	191
Dd	674	WATTATGTGGGAAACMAAATIDMTTTTTMMAAAAATTWKTWTITTDATCTRWATCTKA	733
OY	192	ACAGTTAAGAACAGTCCTTTTGTCTGCACAGAACCCACCACCCTTTCCCAATCACGCC	251
Dd	734	DAATTWGTTRTMTWKATGMTKCKKANMKTAMABRMANMDATRTMAAATWTDTW	793
OY	252	CYTGTGTGCACGAGCCACACAGAGAAGATCCCATGAATGTTAAGGACGTTATCACATG	311
Dd	794	ATAATTCTBRSMNMCAGAAATATSKKKRGAMTBACMKGARBRAKAAAMKAMMATKKK	853
OY	312	TTCAATTCAAACCTTAGCCCTGGTAATGTCCTTAGGCTTAAACGAAACGCTTAAGAG	371
Dd	854	WRADDKKATETPMKSCKGAMKKNWTLMRTKRRTDRTAAAGGMAAANKCATVAAPAA	913

DB	372 AM 373	1 :	914 AM 915
RESULT 5	AV223573	274 bp	mRNA
LOCUS	AV223573	linear	EST 14-NOV-2001
DEFINITION	AV223573 RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue Mus musculus cDNA clone 3830412F22 3', mRNA sequence.		
ACCESSION	AV223573		
VERSION	AV223573.1		
KEYWORDS	EST		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 274)		
AUTHORS	Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kato, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Ode, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shireki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, N., Tomioka, N., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.		
TITLE	RIKEN Mouse ESTs (Kono, H., et al. 1999)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.		
	Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itoh, M., Kikuchi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.		
	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y.		
	High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details.		
FEATURES	Location/Qualifiers		
source	1. .274 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="3830412F22" /clone_id="RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue" /sex="female" /tissue-type="placenta and extra embryonic tissue" /dev_stage="18 days pregnant, adult" /lab_host="DH10B" /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was		

BASE COUNT	82 a	40 c	50 g	102 t
ORIGIN				
Query Match		7.8%	Score 39.2;	DB 9; Length 274;
Best Local Similarity		52.4%;	Pred. No. 1.9;	
Matches	86;	Conservative	0;	Mismatches 78; Indels 0; Gaps 0;
QY	10	CTAAAGCAGCACCAATATTTTCAAAAGCCTCAATCAATTAATTTTCTGATTCATTCGC	69	
DB	27	CCAAATCCACCATGTGATTAAATGCTAGCTGTGAATCTTGATTAAGTAATTTCTT	86	
QY	70	ATTAATTGTGTACATTAATTAATTAATTTGATTTGCTGTGTCCTCCAGTCATCCCAATTAACA	129	
DB	87	ATTTTGTGAATTTTAAATTTTGATGAGGGGGGCTATGAAGTGTAGAGAAATTAAGTTCTC	146	
QY	130	TCGATTTGGCAGCCAAATCCTTATTTGGTGCAATACATATCA	173	
DB	147	CCAAATATTTGCATATATTAGTTTGTGTTGTAACATGCTTTCTTGA	190	
RESULT 6				
LOCUS	BF423025/			
DEFINITION	BF423025	497 bp	mRNA	linear
ACCESSION	Hc_d11_21B10.SKPL	Haemonchus contortus d11 mixed adult from David		
VERSION	BF423025.1	GI:11411014		
KEYWORDS	EST.			
SOURCE	Haemonchus contortus.			
ORGANISM	Haemonchus contortus			
REFERENCE	Eukaryota, Metazoa, Nematoda; Chromadorea; Rhabditida; Strongylida;			
AUTHORS	Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.			
TITLE	1 (bases 1 to 497)			
JOURNAL	Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guillian, D., Hall			
COMMENT	, N., Quayle, M. and Barrett, B.			
	Edinburgh University/Sanger Centre Nematode EST Project			
	Unpublished (2000)			
	Contact: Blaxter ML			
	Institute of Cell, Animal and Population Biology			
	University of Edinburgh			
	Asworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9			
	3J7, UK.			
	Tel: +44 131 650 6760			
	Fax: +44 131 670 5450			
	Email: mark.blaxter@ed.ac.uk			
	The library was prepared by David Knox of the Moredun Institute,			
	Edinburgh. Sequencing was performed by the Pathogen Sequencing Unit,			
	Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart			
	Barrell). The sequence contained a Polya tail (trimmed)			
	PCR Primers			
	FORWARD: T3			
	BACKWARD: T7/PL			
	Plate: 21 row: B column: 10			
	Seq primer: SKPL			
	High quality sequence stop: 497.			
	Location/Qualifiers			
	1. 497			
	/organism="Haemonchus contortus"			
	/strain="Moredun"			
	/db_xref="taxon:6289"			
	/clone="Hc_d11_21B10"			

```

clone.lib-"Haemochus confortus d11 mixed adult from
David Knox"
/dev_stage="Adult, day 11"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI
Constructed by David Knox, Moredun Institute, Edinburgh.
Primary library titre was 4x106 pfu."

BASE COUNT      207 a      30 c      59 g      201 t

ORIGIN
Query Match      7.7%; Score 38.8; DB 12; Length 497;
Best Local Similarity 60.4%; Pred. No. 2.9;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 23 AATATTTCAAAGCTAAGTCAATAATTAATTTACTGATTCATATTCGCATTATGTGGTAC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 AAAATATTAATTAATTTACCCTAATAATTTTATTTTTTAAAAACAATTAATTTTATATA 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 83 ATAAATTGATTTGATTTGCGTGCTCCCAAGGCCATCCCATTAATGCC 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 TTAAACAATAATTAANGTTAAAATTTCGAAGTAGTACCGCATTATTTC 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
BH539949/c
LOCUS          574 bp DNA linear GSS 14-DEC-2001
DEFINITION     BOHH179TF BOHN Brassica oleracea genomic clone BOHH179, DNA
sequence.
ACCESSION      BH539949
VERSION        BH539949.1 GI:17784813
KEYWORDS       GSS.
SOURCE         Brassica oleracea.
ORGANISM       Brassica oleracea.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 574)
AUTHORS        Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE          Whole genome shotgun sequencing of Brassica oleracea
JOURNAL        Unpublished (2001)
COMMENT        Other GSSs: BOHH179TR
                Contact: Chris Town
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel.: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TF
                Class: sheared ends.
FEATURES
source         Location/Qualifiers
                1..574
                 /organism="Brassica oleracea"
                 /strain="N01000DH3"
                 /db_xref="taxon:3712"
                 /clone="BOHH179"
                 /clone="BOHH179"
                 /clone.lib="BOHN"
                 /note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
                 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT      215 a      62 c      79 g      218 t

ORIGIN
Query Match      7.7%; Score 38.8; DB 17; Length 574;
Best Local Similarity 60.4%; Pred. No. 3;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 23 AATATTTCAAAGCTAAGTCAATAATTAATTTACTGATTCATATTCGCATTATGTGGTAC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 AATGATTCAAATGCAATGCTATTTGGTGTTCATAGTATATATTTTATGATTTGTATA 48
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 83 ATAATTTGATTTGATTTGCGTGCTCCCAAGGCCATCCCATTAATGCC 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47 AAAATATGATTTTACTTCGGGTGATGCCAACCTGATTCCTATPACCC 2
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
RESULT 8 B1546960/c		B1546960	60319017551	NIH_MGC_95 Homo sapiens	cdna clone	IMAGE:5261562 5',					
LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
885 bp	mrna	linear	EST 05-SEP-2001								
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches	82;	Conservative	0;	Mismatches	72;	Indels	0;	Gaps
Query Match	7.7%	Score 38.8;	DB 13;	Length 885;							
Best Local Similarity	53.2%	Pred. No. 3.3;	Matches								

[illegible]

	COMMENT	Contact: Tadashu Shin- Center For Genetic Resource Information National Institute of Genetics 1111 Yatae, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshin@genes.nig.ac.jp.
FEATURES	SOURCE	location/Qualifiers
		1..414 /organism="Drosophila melanogaster" /strain="AX4" /db_xref="taxon:44689" /clone="dd332c06" /collection="dictyostelium discoidium cDNA library, SF" /dev_stage="slug stage"
BASE COUNT	188 a	30 c
ORIGIN	21 g	160 t
		15 others
Query Match	7.7%	Score 38.6; DB 13; Length 414;
Best Local Similarity	56.3%	Pred. No. 3.1;
Matches	71; Conservative	0; Mismatched 55; Indels 0; Gaps 0
OY	5 ATGAACATAAGCACCACAATTTTCAAAAGCTAAGTCATAATATTACTTGTAATCAA	64
DB	179 ATTGACATTAATCTCCAAAAATTTCTTAATCATAAAATTAATATATATATATTA	238
OY	65 TTGCATATTGTCTTACATAATAGAATTTGATTTGGTTCCTCAAGCTCCATCAATT	124
DB	239 TTTATGTTTTGNGTGTTAAATTAATATGTTTGGCGCAATCAATTTATTTT	298
OY	125 ATCCAT	130
DB	299 AAAAAT	304
RESULT 11		
LOCUS	BI585460	
DEFINITION	BI585460.5prime RH Drosophila melanogaster normalized Head pfic-1	
ACCESSION	Drosophila melanogaster cdna clone RH25047 5, mRNA sequence.	
VERSION	BI585460.1 GI:15476882	
KEYWORDS	EST.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 641)	
AUTHORS	Stapleton,M., Brocksheip,P., Hong,L., Tyler,D., Berman,B., Carlson, J., Chameje,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George, R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Paley,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celinker,S. and Rubin, G.M.	
TITLE	BGP/HMMI RH Drosophila EST Project	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Stapleton, M.	
	BGP	
	Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu hit genomic AE003431: arm:x [419619,4510082] estimated-cyto:c41-c48: 08/18/2001 Plate: RH.250 row: D column: 11 High quality sequence stop: 516. Location/Qualifiers 1..641 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="RH25047"	
FEATURES	SOURCE	



```

/strain="HM1:IMSS" |
/db.xref="taxon:5759"
/clone.lib="Entamoeba histolytica Sheared DNA"
/note="Vector: Site_1: Bst I; Constructed at The
Institute For Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith.

```

`/db.xref="Caxon:5759"  
/clone.lib="Entameba histolytica Sheared DNA"  
/note="vector: pHS1; Site_1: Bst I; constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entameba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450 ). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The  $\lambda$  +  $\lambda$  method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for`



Wed Jul 9 09:27:38 2003

us-10-083-853b-2\_copy\_18900\_19400.rst

Page 9

BASE COUNT	87 a	74 c	119 g	145 t
ORIGIN	Pieter de Jong			

Query Match	7.6%	Score 38.2;	DB 17;	Length 425;
Best Local Similarity	52.9%	Pred: NO. 4.1;		
Matches 82; Conservative	0;	Mismatches 73;	Indels 0;	Gaps 0;

Oy	138	GACGCAAAATCCTCTTATTGGTGCAACATACATAGTAGTCCTTTCCACAAGAACAAGATT	197
Dd	211	GCACCCAAAGGCCCTAAGTAACCTTGGACAAAGGAAGATTGTTCATCACAAAAACAATGTC	152
Oy	198	AGAACCAAGTCCCTTTGTCTGCACACGAACCACACOCCTTTCGCANTGCACGCCCTTGTC	257
Dd	151	AACGCCCTGTCTATGTTCATTAGCAAATACCTTGAGAAAAGAAATCCCAAAAGTGTACACATTGG	92
Oy	258	TGCACACGCGCCACACAGAGAGAAATCCCATGATG	292
Dd	91	TTCTAGACTCACCCCAGACAGCAACACAAATGGAAG	57

Search completed: July 4, 2003, 17:35:39  
Job time : 1100.73 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 16:40:54 ; Search time 112.075 Seconds

(without alignments)  
6978.003 Million cell updates/sec

File: US-10-083-853b-2\_COPY\_18900\_19400

Perfect score: 501  
Sequence: 1 ttctatgaactaaagcacc.....gacccctgttcttctttt 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Published Applications: NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq: \*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	29921	9	US-10-083-853-1
2	37	7.4	11036	9	US-10-239-676-118
3	36.2	7.2	924	10	US-09-939-980-83
4	36.2	7.2	993	9	US-10-091-504-2307
5	36.2	7.2	993	9	US-10-091-504-2308
6	36.2	7.2	993	10	US-09-764-869-2307
7	36.2	7.2	993	10	US-09-764-869-2308
8	35.8	7.1	2000	9	US-09-938-842A-4432
9	35.4	7.1	952	9	US-10-091-504-547
10	35.4	7.1	952	10	US-09-764-869-547
11	34	6.8	7903	9	US-10-239-676-109
12	33.8	6.7	555	9	US-10-123-155-72
13	33.8	6.7	640681	10	US-09-790-988-1
14	33.6	6.7	1954	9	US-10-103-313-282
15	33.2	6.6	654	9	US-10-239-676-45
16	33.2	6.6	15598	7	US-08-781-986A-82
17	33.2	6.6	31314	10	US-09-764-877-3875
18	33	6.6	1565	9	US-09-938-842A-4527
19	33	6.6	8085	9	US-10-172-086-39

20	32.8	6.5	531	10	US-09-815-242-9316	Sequence 9316, Ap
21	32.8	6.5	531	10	US-09-815-242-9469	Sequence 9469, Ap
22	32.8	6.5	987	12	US-10-044-090-494	Sequence 494, App
23	32.6	6.5	32195	9	US-10-102-627-92	Sequence 92, Appl
24	32.6	6.5	37950	9	US-09-853-526-183	Sequence 183, App
25	32.6	6.5	37950	10	US-09-901-484A-183	Sequence 183, App
26	32.6	6.5	66804	10	US-09-740-041-3	Sequence 3, Appl1
27	32.6	6.5	1830121	9	US-10-329-960-1	Sequence 1, Appl1
28	32.4	6.5	8693	9	US-10-172-086-38	Sequence 38, Appl
29	32.4	6.5	15732	9	US-10-239-676-95	Sequence 95, Appl
30	32.2	6.4	298	10	US-09-864-761-18143	Sequence 18143, A
31	32.2	6.4	496	9	US-09-796-692-4404	Sequence 4404, Ap
32	32.2	6.4	496	9	US-10-040-862-4404	Sequence 4404, Ap
33	32.2	6.4	63588	9	US-10-243-735-3	Sequence 3, Appl1
34	32.2	6.4	358	10	US-09-964-824A-13	Sequence 13, Appl
35	32	6.4	418	10	US-09-983-965-422	Sequence 422, App
36	32	6.4	1031	9	US-10-106-698-1745	Sequence 1745, Ap
37	32	6.4	2000	9	US-09-938-842A-3063	Sequence 3063, Ap
38	32	6.4	45862	9	US-10-216-355-3	Sequence 3, Appl1
39	32	6.4	119596	9	US-10-270-336-3	Sequence 3, Appl1
40	32	6.4	402850	9	US-09-844-891-9832	Sequence 9832, Ap
41	31.8	6.3	532	9	US-09-764-891-9832	Sequence 9832, Ap
42	31.8	6.3	2718	9	US-10-120-687-60	Sequence 60, Appl
43	31.8	6.3	2719	10	US-09-866-866A-9	Sequence 9, Appl1
44	31.8	6.3	2788	10	US-09-745-763-196	Sequence 196, App
45	31.8	6.3	4864	9	US-10-149-640-15	Sequence 15, Appl

#### ALIGNMENTS

RESULT 1  
US-10-083-853-1  
Sequence 1, Application US/10083853  
Patent No. US20020164709A1  
GENERAL INFORMATION:  
APPLICANT: Affymetrix, Inc  
APPLICANT: Shigeta, Ron T  
APPLICANT: Shigeta, Ron T  
TITLE OR INVENTION: Nucleic Acid Encoding Growth Factor Protein  
FILE REFERENCE: 3385.1  
CURRENT APPLICATION NUMBER: US/10/083, 853  
CURRENT FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: USSN 60/272, 663  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 1  
LENGTH: 29921  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-10-083-853-1

Query Match 100.0%; Score 501; DB 9; Length 29921;  
Best Local Similarity 100.0%; Pred. No. 2.9e-131;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTTCATGACCTAAAGCAGCAATATTTTCAAGCTAAGCTAATATTTACTTAAT	60
DB	18900	TTTCATGACCTAAAGCAGCAATATTTTCAAGCTAAGCTAATATTTACTTAAT	18959
QY	61	TCATTCGCAATATTTGTGTACATTAATGATTTGTTTCGTTCCAAAGCCATTC	120
DB	18960	TCATTCGCAATATTTGTGTACATTAATGATTTGTTTCGTTCCAAAGCCATTC	19019
QY	121	AATTCATCATTCATTTGCGAGCAATATTTGTCATCATATCATATGATATCTT	180
DB	19020	AATTCATCATTCATTTGCGAGCAATATTTGTCATCATATCATATGATATCTT	19079
QY	181	CTCAGCAAGACAGCTTAGAAGCAAGCTCTTTGTCTGACAGAACCCACCCCTTCC	240
DB	19080	CTCAGCAAGACAGCTTAGAAGCAAGCTCTTTGTCTGACAGAACCCACCCCTTCC	19139

QY 241 CAATGACGCCCTTGTGTGACACGACACAGAAAGTATCCATGATTTAGGAC 300  
DB 19140 CAATGACGCCCTTGTGTGACACGACACAGAAAGTATCCATGATTTAGGAC 19199  
QY 301 TTATCACCATTGTTCACTTAAGCTTAGCCCTGTAAATGCTTAGGCTTAAGCTGAAC 360  
DB 19200 TTATCACCATTGTTCACTTAAGCTTAGCCCTGTAAATGCTTAGGCTTAAGCTGAAC 19259  
QY 361 TCGTTAAGGAGAAAGAAAAAGCCTTTGGACACAGTCAAAAGTCAAGACAGT 420  
DB 19260 TCGTTAAGGAGAAAGAAAAAGCCTTTGGACACAGTCAAAAGTCAAGACAGT 19319  
QY 421 GTCAGTAGAATGATCTTAAGAAATTCAGACAAAGTGTGACCATGATCCACTAAC 480  
DB 19320 GTCAGTAGAATGATCTTAAGAAATTCAGACAAAGTGTGACCATGATCCACTAAC 19379  
QY 481 CGCATCCCTGTTGTCTTTT 501  
DB 19380 CGCATCCCTGTTGTCTTTT 19400

## RESULT 2

US-10-239-676-118  
; Sequence 118, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10033529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 118  
; LENGTH: 11036  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (9056, 9079, 9081, 9094, 9108..)  
US-10-239-676-118

Query Match 7.4%; Score 37; DB 9; Length 11036;  
Best Local Similarity 60.4%; Pred. No. 9;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 23 AATATTTCAAGCTAAGTCATTAATATTACTGTAATTCATTCGATTTATTTGTTAC 82  
DB 4240 AATTTTGACAAAGTTTGTAGTATTATTATTAATAGTAATCATATTATTATTATTA 4299  
QY 83 ATAAATTAGATTGATTTCGTGTCCTCAAGTCATCAAT 123  
DB 4300 ATTAATATTTCGATTTTATTATTATTAGTGTTAAT 4340

## RESULT 3

US-09-939-980-83/C  
; Sequence 83, Application US/09939980  
; Patent No. US20020082234A1  
; GENERAL INFORMATION:

APPLICANT: Black, Michael  
Burnham, Martin  
Hodgson, John  
Knowles, David  
Lometto, Michael  
Nicholas, Richard  
Pratt, Julie  
Reichard, Richard  
Rosenberg, Martin  
Ward, Judith  
TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,  
Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,980  
FILING DATE: 27-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/936,165  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 924 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 83:  
US-09-939-980-83  
Query Match 7.2%; Score 36.2; DB 10; Length 924;  
Best Local Similarity 51.7%; Pred. No. 4;  
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
QY 25 TATTTTCAAGCTAAGTCATTAATATTACTGTAATTCATTCGATTTATTTGTTACT 84  
DB 536 TATTTTGAATCTTTATACCAATTAATCTAGACTATNNNGCATTAATTATATGATAT 477  
QY 85 AATATAGATTGATTTCGTGTCCTCAAGTCATCAATTAATTCATTCGATTTATTTG 144  
DB 476 CAAATGCGCTNNATTCCTTGTCTTNNATGCATCTACATCTTTAGNANTTNGNAGNCA 417

## RESULT 4

US-10-091-504-2307/C  
; Sequence 2307, Application US/10091504  
; Publication No. US20030059908A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC0007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07

NUMBER OF SEQ ID NOS: 2442  
Prior Application removed - See File Wrapper or Palm  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 2307  
LENGTH: 993  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-091-504-2307

Query Match  
Best Local Similarity 53.18; Pred. No. 4.1;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTCAAGCTAGTCAATAATTTACTGATTCATTCGCGATTATGTTAC 82  
DB 677 AAAAAATGAAAAAATACATCACTTAGTCTAATAGAAATATAGTACAGTGAGAG 618  
QY 83 AATAATTAGATTGATTGCTGTTCTCCAAAGTCATCCATTCATTCATTGGCAGC 142  
DB 617 AGAAAGATTATATCAGAGATGCTCTTAAGTCATCTCATTTGTTTCAATATAATG 558  
QY 143 CAATCCCTTATGTCGATACAT 167  
DB 557 TACATCTGATTACATATACAAACAT 533

RESULT 5  
US-10-091-504-2308/c  
Sequence 2308, Application US/10091504  
Publication No. US2003005908A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC007C1  
CURRENT APPLICATION NUMBER: US/10/091,504  
CURRENT FILING DATE: 2002-03-07  
NUMBER OF SEQ ID NOS: 2442  
Prior Application removed - See File Wrapper or Palm  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 2308  
LENGTH: 993  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-091-504-2308

Query Match  
Best Local Similarity 53.18; Pred. No. 4.1;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTCAAGCTAGTCAATAATTTACTGATTCATTCGCGATTATGTTAC 82  
DB 677 AAAAAATGAAAAAATACATCACTTAGTCTAATAGAAATATAGTACAGTGAGAG 618  
QY 83 AATAATTAGATTGATTGCTGTTCTCCAAAGTCATCCATTCATTCATTGGCAGC 142  
DB 617 AGAAAGATTATATCAGAGATGCTCTTAAGTCATCTCATTTGTTTCAATATAATG 558  
QY 143 CAATCCCTTATGTCGATACAT 167  
DB 557 TACATCTGATTACATATACAAACAT 533

RESULT 6  
US-09-764-869-2307/c  
Sequence 2307, Application US/09764869  
Patent No. US20020061521A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC007  
CURRENT APPLICATION NUMBER: US/09/764,869  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 2442  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 2307  
LENGTH: 993  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-869-2307

Query Match  
Best Local Similarity 53.18; Pred. No. 4.1;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTCAAGCTAGTCAATAATTTACTGATTCATTCGCGATTATGTTAC 82  
DB 677 AAAAAATGAAAAAATACATCACTTAGTCTAATAGAAATATAGTACAGTGAGAG 618  
QY 83 AATAATTAGATTGATTGCTGTTCTCCAAAGTCATCCATTCATTCATTGGCAGC 142  
DB 617 AGAAAGATTATATCAGAGATGCTCTTAAGTCATCTCATTTGTTTCAATATAATG 558  
QY 143 CAATCCCTTATGTCGATACAT 167  
DB 557 TACATCTGATTACATATACAAACAT 533

RESULT 7  
US-09-764-869-2308/c  
Sequence 2308, Application US/09764869  
Patent No. US20020061521A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC007  
CURRENT APPLICATION NUMBER: US/09/764,869  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 2442  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 2308  
LENGTH: 993  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-869-2308

Query Match  
Best Local Similarity 53.18; Pred. No. 4.1;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTCAAGCTAGTCAATAATTTACTGATTCATTCGCGATTATGTTAC 82  
DB 677 AAAAAATGAAAAAATACATCACTTAGTCTAATAGAAATATAGTACAGTGAGAG 618  
QY 83 AATAATTAGATTGATTGCTGTTCTCCAAAGTCATCCATTCATTCATTGGCAGC 142  
DB 617 AGAAAGATTATATCAGAGATGCTCTTAAGTCATCTCATTTGTTTCAATATAATG 558  
QY 143 CAATCCCTTATGTCGATACAT 167  
DB 557 TACATCTGATTACATATACAAACAT 533

RESULT 8  
US-09-938-842A-4432/c  
Sequence 4432, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kieps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SRIPI300-3



;; CURRENT APPLICATION NUMBER: US/09/938,842A  
;; CURRENT FILING DATE: 2001-08-24  
;; PRIOR APPLICATION NUMBER: US 60/227,866  
;; PRIOR FILING DATE: 2000-08-24  
;; PRIOR APPLICATION NUMBER: US 60/264,647  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: US 60/300,111  
;; PRIOR FILING DATE: 2001-06-22  
;; NUMBER OF SEQ ID NOS: 5379  
;; SEQ ID NO: 4432  
;; LENGTH: 2000  
;; TYPE: DNA  
;; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4432

Query Match  
Best Local Similarity 54.3%; Pred. No. 7.8;  
Matches 94; Conservative 0; Mismatches 77; Indels 2; Gaps 1;

QY 8 AACCTAAAGCACCACAAATATTTTCAAGCTAAAGTCAATTAATTTAGTGAATTCATTC 67  
DB 1459 AATTAATGTAATCTCTTCAATTTCAATAGTTAATGATTAATTAATTAATTAATTT 1400  
QY 68 GCATTTATGTTGTTACATAATAGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 127  
DB 1399 TACATATTTATTTT--ATTATACATATTTAATGTTGTTGTTGTTGTTGTTGTTGTT 1342  
QY 128 CATCCATTTGTCAGCCAAATCCTTATGTTGTCATACATACATGATTCCTT 180  
DB 1341 ACATGATTTTATTAATTAATCAATTTTAATAGTTTCTCATCTCGAAGATTTT 1289

RESULT 9  
US-10-091-504-547/c  
; Sequence 547, Application US/10091504  
; Publication No. US20030059908A1

;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC007C1  
;; CURRENT APPLICATION NUMBER: US/10/091,504  
;; CURRENT FILING DATE: 2002-03-07  
;; NUMBER OF SEQ ID NOS: 2442  
;; Prior Application removed - See File Wrapper or Palm  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO: 547  
;; LENGTH: 952  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (35)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc\_feature  
;; LOCATION: (385)  
;; OTHER INFORMATION: n equals a,t,g, or c  
US-10-091-504-547.

Query Match  
Best Local Similarity 51.7%; Pred. No. 6.8;  
Matches 75; Conservative 2; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTTCAAGCTAAAGTCAATTAATTTAGTGAATTCATTAATTCGATTTGTTAC 82  
DB 678 AAAAAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 619  
QY 83 AATATTTAGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 142  
DB 618 AGAAAGTATTTATCGAGATGCTCTTTAAGTCATTCATCTGTTTTCMAATATAAATG 559  
QY 143 CAATTCCTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 167  
DB 558 TACATCTGATTACATATTCATATTCAT 534

RESULT 10  
US-09-764-869-547/c  
; Sequence 547, Application US/09764869  
; Patent No. US20020061521A1

;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC007  
;; CURRENT APPLICATION NUMBER: US/09/764,869  
;; Prior application data removed - refer to PALM or file wrapper  
;; NUMBER OF SEQ ID NOS: 2442  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO: 547  
;; LENGTH: 952  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens

Query Match  
Best Local Similarity 51.7%; Pred. No. 6.8;  
Matches 75; Conservative 2; Mismatches 68; Indels 0; Gaps 0;

QY 23 AATATTTTCAAGCTAAAGTCAATTAATTTAGTGAATTCATTAATTCGATTTGTTAC 82  
DB 678 AAAAAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 619  
QY 83 AATATTTAGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 142  
DB 618 AGAAAGTATTTATCGAGATGCTCTTTAAGTCATTCATCTGTTTTCMAATATAAATG 559  
QY 143 CAATTCCTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 167  
DB 558 TACATCTGATTACATATTCATATTCAT 534

RESULT 11  
US-10-239-676-109/c  
; Sequence 109, Application US/10239676  
; Publication No. US20030082609A1

;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENBROCK, Christian  
;; APPLICANT: BERLIN, Kurt  
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
;; FILE REFERENCE: 5013.1003  
;; CURRENT APPLICATION NUMBER: US/10/239,676  
;; CURRENT FILING DATE: 2002-09-24  
;; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
;; DE 10019058.8  
;; DE 10019173.8  
;; DE 10032529.7  
;; DE 10043826.1

;; PRIOR FILING DATE: 2001-04-06

;; 2000-04-06  
;; 2000-04-07  
;; 2000-06-30  
;; 2000-09-01  
;; NUMBER OF SEQ ID NOS: 228  
;; SEQ ID NO: 109  
;; LENGTH: 7903  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:

```

; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12

```

```

: NUMBER OF SEQ ID NOS: 653
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 283
: LENGTH: 1954
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: (1598)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc-feature
: LOCATION: (1599)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc-feature
: LOCATION: (1907)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc-feature
: LOCATION: (1915)
: OTHER INFORMATION: n equals a,t,g, or c
: US-10-103-313-282

```

Query Match	6.7%	Score	33.6;	DB	9;	Length	1954;
Best Local Similarity	63.8%;	Pred. No.	33;				
Matches	51;	Conservative	0;	Mismatches	29;	Indels	0;
						Gaps	0;

QY	89	TAGATTTCATTTGCGTTCCTCCAAATCCATCCAAATATCCATTCATTGGACGCCAAATC	148
Db	675	TAGCGCCATTTCAAGTGAGTCTCTACCCCGAGGATTAATCATTCATTCATTGTGAACGAATTC	616
QY	149	CTCTTATGTGGTCATACATA	168
Db	615	CTCTCATATAGTTTCATTAAACA	596

```

RESULT 15
US-10-239-676-45/c
; Sequence 45, Application US/10239676
; Publication NO. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239.676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 45
; LENGTH: 6544
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-45

```

Query Match	6.68;	Score 33.2;	DB 9;	Length 6544;
Best Local Similarity	48.98;	Pred. No. 82;		
Matches 89; Conservative	0;	Mismatches 93;	Indels 0;	Gaps 0;

QY 2TCGATGACCTAAAGCACCACCAATATTTTCAACCTACGCTAATATATTTACTGGAATT 61  
Db 1364 TCGATATTTTAATTTCTCCACCAATTCACATCTAAGCTTAATTCACATCATTTATATATCCCATC 1305

Oy	62	CAATTCGGCATTAATGTGTACATTAATAGATTTGATTCGTCTTCGCAAGTCGATCGCA	121
Db	1304	TAACTCTATATACCATTTTACTTTTTTAATTAATTTACTCAAAATTCACCAATTTCCATAT	1245
Oy	122	ATTATTCATCCATTTGGCAGCCAAATCCCTATTATGGCGCTACATATCATGATATCTTC	181
Db	1244	ATTAACATATATCATCTTCTATATAAAAAATAATTAATTAACCAATTAATAATTCAAATATCTTC	1185
Oy	182	TC 183	
Db	1184	TC 1183	

Search completed: July 4, 2003, 19:36:13  
Job time : 114.075 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:16:37; Search time 35.0233 seconds

(without alignments)  
4386.938 Million cell updates/sec

File: us-10-083-853b-2 copy\_18900\_19400  
Perfect score: 501  
Sequence: 1 tctatgaactaaagcacc.....gcattccgtttgtctttt 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database 8 Issued: Patents NMAO

1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.2	7.2	924	US-08-936-165A-83	Sequence 83, Appl
2	35.8	7.1	3030	US-09-693-146-3	Sequence 3, Appl
3	33	6.6	6407	US-07-977-630-1	Sequence 1, Appl
4	33	6.6	6407	US-07-977-630-2	Sequence 2, Appl
5	33	6.6	7218	US-08-232-463-14	Sequence 14, Appl
6	32.8	6.5	5558	US-08-961-527-103	Sequence 103, Appl
7	32.6	6.5	37950	US-09-338-907-183	Sequence 183, Appl
8	32.6	6.5	37950	US-09-218-207-183	Sequence 183, Appl
9	32.4	6.4	811	US-09-662-250A-15	Sequence 15, Appl
10	32.2	6.4	188	US-08-330-108-2	Sequence 2, Appl
11	32.2	6.4	188	PCT-US92-10087-2	Sequence 2, Appl
12	32.2	6.4	731	US-08-451-105A-2	Sequence 2, Appl
13	32.2	6.4	1223	US-08-463-081B-3	Sequence 3, Appl
14	32.2	6.4	1223	US-08-461-379A-3	Sequence 3, Appl
15	32.2	6.4	1223	US-08-462-390B-3	Sequence 3, Appl
16	32.2	6.4	1223	US-08-463-074B-3	Sequence 3, Appl
17	32.2	6.4	1223	US-08-465-585C-3	Sequence 3, Appl
18	32.2	6.4	1223	US-08-652-446-3	Sequence 3, Appl
19	32.2	6.4	1278	US-09-134-001C-2817	Sequence 2817, Appl
20	31.8	6.3	4915	US-09-173-053-7	Sequence 7, Appl
21	31.8	6.3	5676	US-08-663-998-3	Sequence 3, Appl
22	31.8	6.3	5676	US-08-663-998-4	Sequence 4, Appl
23	31.8	6.3	5845	US-09-173-053-1	Sequence 1, Appl
24	31.8	6.3	5899	US-09-173-053-2	Sequence 2, Appl
25	31.8	6.3	5900	US-08-663-998-1	Sequence 1, Appl
26	31.8	6.3	5952	US-08-663-998-2	Sequence 2, Appl
27	31.6	6.3	353	US-09-366-887A-21	Sequence 21, Appl

28	31.6	6.3	605	US-09-366-887A-26
29	31.6	6.3	44453	US-09-146-053-5
30	31.2	6.2	1695	US-09-134-001C-2302
31	31.2	6.2	3301	US-08-447-430A-42
32	31.2	6.2	3423	US-08-447-430A-40
33	31.2	6.2	3423	US-09-485-737B-86
34	31.2	6.2	3474	US-08-447-430A-41
35	31.2	6.2	3474	US-08-318-837-10
36	31.2	6.2	4009	US-08-500-860A-2
37	31	6.2	282	US-09-134-001C-2050
38	31	6.2	918	US-09-134-001C-355
39	31	6.2	3892	US-08-569-214-1
40	31	6.2	3892	US-08-937-236-1
41	30.8	6.1	3474	US-08-122-458D-8
42	30.4	6.1	713	US-08-385-335A-10
43	30.4	6.1	6285	US-08-659-206A-1
44	30.4	6.1	7287	US-08-659-206A-1
45	30.2	6.0	854	US-08-998-416-534

#### ALIGNMENTS

RESULT 1  
US-08-936-165A-83/c  
Sequence 83, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Piratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027, 032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ. ID NO.: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 924 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

Sequence 26, Appl  
Sequence 5, Appl  
Sequence 2302, Ap  
Sequence 42, Appl  
Sequence 40, Appl  
Sequence 86, Appl  
Sequence 41, Appl  
Sequence 10, Appl  
Sequence 2, Appl  
Sequence 2050, Ap  
Sequence 355, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 8, Appl  
Sequence 10, Appl  
Sequence 4, Appl  
Sequence 1, Appl  
Sequence 534, App

TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-936-165A-83

Query Match 7.2%; Score 36.2; DB 4; Length 924;  
Best Local Similarity 51.7%; Pred. No. 0.22; Indels 0; Gaps 0;  
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 25 TATTTTCAAGCTAAGTCAATAATTTTCAAGTCAATTCGCAATTTGTTTACAT 84  
DB 536 TATTTGATCTTTTACCAATTAATCTAGACATANNNGCATATTNATGATAT 477  
QY 85 AATTAAGATTGATTTGCTGTTCTCCAGTCCCAATTAATTCATTCATTGGCAGCCA 144  
DB 476 CAAATGCGCTNNATTCCTGCTTNNMATGCATCTACATCTTAGNANTTNGNAGNCA 417

## RESULT 2

US-09-693-146-3/c  
Sequence 3, Application US/09693146  
Patent No. 6413758  
GENERAL INFORMATION:  
APPLICANT: Xu, Shuang-Yong  
APPLICANT: Xiao, Jian-ping  
APPLICANT: Zhu, Zhenyu  
TITLE OF INVENTION: Method For Cloning And Expression Of Bpmi Restriction  
FILE REFERENCE: NEB-183  
CURRENT APPLICATION NUMBER: US/09/693,146  
CURRENT FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 3030  
TYPE: DNA  
ORGANISM: Bacillus pumilus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3030)  
US-09-693-146-3

Query Match 7.1%; Score 35.8; DB 4; Length 3030;  
Best Local Similarity 52.3%; Pred. No. 0.45; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 4 TATGACTTAAAGCACACATATTTTCAAGTCAATTAATTAATTTACTTGAATCA 63  
DB 1339 TAACATCAACACCAAAATTAATTTCTCCAGATATTTCTTTGCTTAAAGTCAAAA 1280  
QY 64 ATTGCATTAATTTGTTTACATAAATTAATTAATTTGATTTGCTTCCAGTCCATCAAT 123  
DB 1279 TTAGCCCTTCTTCATTTCACTTAATTAATTCGAATCATCGATGTTCTCTTAATTAATCT 1220  
QY 124 TATCATCATTTGGCAGCAAAATCTCTTA 154  
DB 1219 TTTCATTAATTTCTCTACTAGAAATCATTA 1189

## RESULT 3

US-07-977-630-1  
Sequence 1, Application US/07977630  
Patent No. 5583038

GENERAL INFORMATION:  
APPLICANT: Slover, Charles K.  
TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING  
TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS  
NUMBER OF SEQUENCES: 84  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Carella, Byrne, Bain, Giffillan, Cecchi,  
ADDRESSER: Stewart & Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: New Jersey

COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM

OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/977,630  
FILING DATE: No. 5583038ember 17, 1993

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Herion, Charles J.

REGISTRATION NUMBER: 28, 019  
REFERENCE/DOCKET NUMBER: 469201-174

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6407 nucleotides

TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: circular  
MOLECULE TYPE: nucleic acid

US-07-977-630-1

Query Match 6.6%; Score 33; DB 1; Length 6407;  
Best Local Similarity 58.8%; Pred. No. 4; Indels 0; Gaps 0;

Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 103 TGTTCGCAAGTCAATTCATTAATTCATTTGGCAGCAAAATCTTATTGGTCA 162  
DB 4784 TCTTACCTAGATCCCTTTATTAATGAAGCATTTATCAGGTTATGCTCATGAGCGGA 4843  
QY 163 TACATACATGATATCTTCTCCAGAACACAGTAG 199  
DB 4844 TACATATTTGAATGATTTAGAAAATTAACAAATAG 4880

## RESULT 4

US-07-977-630-2/c  
Sequence 2, Application US/07977630  
Patent No. 5583038

GENERAL INFORMATION:  
APPLICANT: Slover, Charles K.

TITLE OF INVENTION: BACTERIAL EXPRESSION VECTORS CONTAINING  
TITLE OF INVENTION: DNA ENCODING SECRETION SIGNALS OF LIPOPROTEINS  
NUMBER OF SEQUENCES: 84

CORRESPONDENCE ADDRESS:  
ADDRESSER: Carella, Byrne, Bain, Giffillan, Cecchi,  
ADDRESSER: Stewart & Olstein  
STREET: 6 Becker Farm Road

CITY: Roseland  
STATE: New Jersey  
COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM

OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/977,630  
FILING DATE: No. 5583038ember 17, 1993

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Herion, Charles J.

REGISTRATION NUMBER: 28, 019  
REFERENCE/DOCKET NUMBER: 469201-174

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

RESULT 7  
S-09-338-907-183  
Sequence 183, Application US/09338907

```
Patent No. 6265546
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CPICP
CURRENT APPLICATION NUMBER: US/09/338,907
EARLIER FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 183
LENGTH: 37950
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: exon
LOCATION: 5259..5328
OTHER INFORMATION: exon2
FEATURE:
NAME/KEY: exon
LOCATION: 12675..12791
OTHER INFORMATION: exon3
FEATURE:
NAME/KEY: exon
LOCATION: 14621..14710
OTHER INFORMATION: exon4
FEATURE:
NAME/KEY: exon
LOCATION: 19822..19912
OTHER INFORMATION: exon5
FEATURE:
NAME/KEY: exon
LOCATION: 21789..21950
OTHER INFORMATION: exon7
FEATURE:
NAME/KEY: exon
LOCATION: 23387..23510
OTHER INFORMATION: exon6
FEATURE:
NAME/KEY: exon
LOCATION: 25520..26016
OTHER INFORMATION: exon8
US-09-338-907-183

Query Match
Best Local Similarity 57.3%; Score 32.6; DB 4; Length 37950;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 28 TTTCAGGCTAGCTAATAATATTACTGTAATTCGATTAATGTTGTTACATAA 87
Db 5889 TTAAATAATTAATCTTAATATTAATTAAGTAAGTAAGTAAGCTTTGTTTAAACATAA 5948
OY 88 TTGATTTGATTTTCGCTGCTCTCCAAAGTCATCCATCAATTATCCAT 130
Db 5949 TTTGCTGTGATTTGAATCTGCGCTTGCATTTTATTACTT 5991
```

RESULT 8  
US-09-218-207-183  
Sequence 183, Application US/09218207  
Patent No. 6346381  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov

```
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: Prostate cancer gene
FILE REFERENCE: GENSET.018CP1
CURRENT APPLICATION NUMBER: US/09/218,207
CURRENT FILING DATE: 1998-12-22
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 183
LENGTH: 37950
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: exon
LOCATION: 5259..5328
OTHER INFORMATION: exon2
FEATURE:
NAME/KEY: exon
LOCATION: 12675..12791
OTHER INFORMATION: exon3
FEATURE:
NAME/KEY: exon
LOCATION: 14621..14710
OTHER INFORMATION: exon4
FEATURE:
NAME/KEY: exon
LOCATION: 19822..19912
OTHER INFORMATION: exon5
FEATURE:
NAME/KEY: exon
LOCATION: 21789..21950
OTHER INFORMATION: exon7
FEATURE:
NAME/KEY: exon
LOCATION: 23387..23510
OTHER INFORMATION: exon6
FEATURE:
NAME/KEY: exon
LOCATION: 25520..26016
OTHER INFORMATION: exon8
US-09-218-207-183

Query Match
Best Local Similarity 57.3%; Score 32.6; DB 4; Length 37950;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 28 TTTCAGGCTAGCTAATAATATTACTGTAATTCGATTAATGTTGTTACATAA 87
Db 5889 TTAAATAATTAATCTTAATATTAATTAAGTAAGTAAGTAAGCTTTGTTTAAACATAA 5948
OY 88 TTGATTTGATTTTCGCTGCTCTCCAAAGTCATCCATCAATTATCCAT 130
Db 5949 TTTGCTGTGATTTGAATCTGCGCTTGCATTTTATTACTT 5991
```

RESULT 9  
US-09-662-250A-15/C  
Sequence 15, Application US/09662250A  
Patent No. 6368856  
GENERAL INFORMATION:  
APPLICANT: Brett P. Monia  
APPLICANT: Jacqueline Wyatt  
TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHORYLASE KINASE BETA EXPRESS  
FILE REFERENCE: R1S-0129  
CURRENT APPLICATION NUMBER: US/09/662,250A  
CURRENT FILING DATE: 2000-09-14  
NUMBER OF SEQ ID NOS: 102  
SEQ ID NO 15  
LENGTH: 811  
TYPE: DNA



```

; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (45)..(100)
; NAME/KEY: CDS
; LOCATION: (609)..(702)
;
US-09-662-250A-15

```

Query Match	6.58;	Score	32.4;	DB	4;	Length	811;
Best Local Similarity	58.28;	Pred.	No..2.7;				
Matches	57;	Conservative	0;	Mismatches	41;	Indels	0;
						Gaps	0;

QY	1	TTTATGACATAAAGCAGCACAATATTTTCAAGCTAAGTCATATAATTACTGTGAT	60
Db	356	TCTCGAAGTACAACCCAGCGACACGCGACGCGACGCTGTCAAGCTTTAATATTCACCTCAAC	297
QY	61	TCAATTCGCATTATGTGTTACATATAATTAGATTGAT	98
Db	296	TAAACAAATTTTATTTTAAACGCGAAAGAAATATGCT	259

```

OS-08-330-108-2/c
Sequence 2, Application US/08330108
Patent No. 5795752
GENERAL INFORMATION:
APPLICANT: Smith, Kendall A.
TITLE OF INVENTION: IL-2-stimulated Gene
Expression
TITLE OF INVENTION: Expression
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,108
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/104,736
FILING DATE:
APPLICATION NUMBER: US/07/796,066
FILING DATE:
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 base pairs
TYPE: nucleic acid
STRANDEDNESS: single stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: human
CELL TYPE: T-cell blast
IMMEDIATE SOURCE:
LIBRARY:
CLONE: 1F5
OS-08-330-108-2

```

Query Match	6.4%	Score 32.2;	DB 1;	length 188;
Best Local Similarity	51.8%;	Pred. No. 1.7;		
Matches 73;	Conservative 0;	Mismatches 60;	Indels 0;	Gaps 0;
55	TTGAATTCATTCGCATTATGTGTGCATATAATTAGTTTGGTTTCTGTTCCCAAGT			114

[illegible]

```

RESULT 1
PCT-US92-10087-2/c
Sequence 2, Application PC/TUS9210087
GENERAL INFORMATION:
APPLICANT: Smith, Kendall A.
TITLE OF INVENTION: IL-2-Stimulated
TITLE OF INVENTION: Expression
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lalive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10087
FILING DATE: 19921118
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/796,066
FILING DATE: 20-NOVEMBER-1991
ATTORNEY/AGENT INFORMATION:
NAME: Decont, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-028PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEetical: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: human
CELL TYPE: T-cell blast
IMMEDIATE SOURCE:
LIBRARY: 1F5
CLONE: 1F5
PCT-US92-10087-2

```

[illegible]







GenCore version 5.1.6  
Copyright (c) 1993-2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:04:06 ; Search time 166.778 seconds  
(without alignments)  
6765.003 Million cell updates/sec

File: JUS-10-083-853b-2\_COPY\_18900\_19400.rng

Perfect score: 501  
Sequence: 1 ttctatgaactaaagacac.....gacccctgttgcctttt 501

Scoring table:  
IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database 8  
N: Geneseq 10100280  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	38.8	7.7	593	ABV52720
2	37.4	7.5	3212	ABL06380
3	37	7.4	11036	AA545411
4	37	7.4	11036	ABK28264
5	36.4	7.3	7441	ABK40057
6	36.4	7.3	7479	AA563344
7	36.4	7.3	2155561	ABN71527
8	36.2	7.2	924	AA783941
9	36.2	7.2	924	AAV53383

C 10	36.2	7.2	993	22	AA536807
C 11	36.2	7.2	993	22	AA536808
C 12	36	7.2	26636	23	ABL16650
C 13	35.8	7.1	3030	24	AAK99724
C 14	35.4	7.1	3030	22	AA535662
C 15	35.4	7.1	3030	22	AAH54882
C 16	35.4	7.1	3734	22	AAH54739
C 17	35.2	7.0	5105	13	AAQ20536
C 18	35.2	7.0	11172	24	ABL34053
C 19	35.2	7.0	14041	22	AAH48024
C 20	35.2	7.0	50000	21	AA664140
C 21	35	7.0	1334	21	AA593310
C 22	35	7.0	1797	21	AA593331
C 23	35	7.0	3823	21	AA593329
C 24	35	7.0	6270	24	ABL33872
C 25	35	7.0	6658	18	AA798882
C 26	35	7.0	6700	24	ABN80305
C 27	34.8	6.9	6559	23	ABL04876
C 28	34.8	6.9	7059	15	AAQ44750
C 29	34.8	6.9	9859	24	ABL34133
C 30	34.6	6.9	220895	24	ABK84798
C 31	34.4	6.9	374	23	ABV52305
C 32	34.4	6.9	542	23	ABV52528
C 33	34.4	6.9	1365	22	AA588109
C 34	34.4	6.9	11963	23	ABL05864
C 35	34.2	6.8	6447	24	ABO66973
C 36	34.2	6.8	6447	24	ABL70547
C 37	34.2	6.8	6447	24	ABL33086
C 38	34.2	6.8	6447	24	ABL34532
C 39	34	6.8	5204	24	ABL32900
C 40	34	6.8	7903	22	AA545402
C 41	34	6.8	7903	22	ABL33362
C 42	34	6.8	7903	24	ABK28247
C 43	34	6.8	12177	24	ABL32651
C 44	34	6.8	611590	21	AAV22303
C 45	34	6.8	1664976	19	AAV21209

#### ALIGNMENTS

RESULT 1	ABV52720	standard; cDNA; 593 Bp.
ID	ABV52720	
AC	ABV52720;	
XX		
DT	17-SEP-2002 (first entry)	
DE	Human prostate expression marker CDNA 52711.	
XX		
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	
KW	pharmacogenomic marker; gene; ss.	
OS	Homo sapiens.	
XX		
PN	W0200160860-A2.	
XX		
PD	23-AUG-2001.	
XX		
PF	20-FEB-2001; 2001WO-US05171.	
XX		
PR	17-FEB-2000; 2000US-183119P.	
PR	16-MAR-2000; 2000US-189862P.	
PR	25-MAY-2000; 2000US-207454P.	
PR	09-JUN-2000; 2000US-211314P.	
PR	18-JUL-2000; 2000US-219007P.	
PR	13-DEC-2000; 2000US-255281P.	
XX		
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX		
XX	Schlegel R, Endege WO, Monahan JE;	
PI		

Human cardiovascular  
Human cardiovascular  
Drosophila melanog  
DNA of the Bpm1 en  
Human cardiovascul  
S. epidermidis gen  
S. epidermidis gen  
Glutamine 2-oxo-gl  
Human immune syste  
Internal control B  
Nucleotide sequenc  
Nucleotide sequenc  
Nucleotide sequenc  
Nucleotide sequenc  
Human immune syste  
Tomato immunity 2  
Human chemically m  
Drosophila melanog  
Soybean lipoxigena  
Human immune syste  
Human immune syste  
Human cDNA differe  
Human prostate exp  
Human prostate exp  
Human FLEXHT-40 nu  
Drosophila melanog  
Human angiogenesis  
Chemically treated  
Human immune syste  
Human metastasis a  
Human immune syste  
Chemically pretrea  
Human immune syste  
DNA transcription  
Human immune syste  
Arabidopsis thalia  
Methanococcus jann

DR WPI: 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer  
 XX  
 XX Claim 1; Page 10221; 11750pp; English.  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 593 BP; 198 A; 79 C; 100 G; 216 T; 0 other;  
 Query Match 7.7%; Score 38.8; DB 23; Length 593;  
 Best Local Similarity 58.8%; Pred. No. 0.68;  
 Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
 QY 1 TTCTGACATAAAGACACCAATATTTCAAGCTAATTAATTTACTGAT 60  
 DB 352 TTCTATGATATTTAAAAATATACATTTATGTATATATATATTTATTTCTAT 411  
 QY 61 TCAATTCGATTTATGTTACATTAATTTGATTTGATTCGTTCCACGT 114  
 DB 412 GTATTTCTATATATTTATTTATATATATTTAAATTTGGGTTCCGAGTTCTAAGT 465  
 RESULT 2  
 ABL06380/c  
 ID ABL06380 standard; cDNA; 3212 BP.  
 XX  
 AC ABL06380;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13622.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PD 27-SEP-2001.  
 PF 23-MAR-2001; 2001MO-US09231.  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI: 2001-656860/75.  
 DR P-PSDB; ABB62277.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX  
 PS Claim 1; SEQ ID NO 13622; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3212 BP; 817 A; 686 C; 870 G; 839 T; 0 other;  
 Query Match 7.5%; Score 37.4; DB 23; Length 3212;  
 Best Local Similarity 52.2%; Pred. No. 2.9;  
 Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
 QY 12 AAAAGCACCAATTTTCAAGCTAATTAATTTACTTGATTCATTCGAT 71  
 DB 183 AAAATCTTATACCACTTTCAAGATTTAGGCCAAATTTATAGATTTCTTACACTT 124  
 QY 72 TATTTGTTACATTAATTTAGATTTGTTGTTGTTCCAGTCATTCATTCATC 131  
 DB 123 TATGTTAGCCATTAACCTGCGTCAAAATTTGCTTGCAGAACCCCTTAATTTCTTC 64  
 QY 132 CATTTGGACGCCAAATTCCTTTATTTGTTGATCATACATA 170  
 DB 63 GCTTTTACGTTATTCATTTACGTAGCTGCCAAATTTCA 25  
 RESULT 3  
 AAS45411  
 ID AAS45411 standard; DNA; 11036 BP.  
 XX  
 AC AAS45411;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Chemically pretreated complementary DNA associated with cell cycle #58.  
 XX  
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KW graft-versus-host disease; glomerular disease; lewy body disease; cancer;  
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KW immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;  
 KW PCR primer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168911-A2.  
 PD 20-SEP-2001.  
 PF 15-MAR-2001; 2001MO-EP02945.  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI: 2001-602751/68.  
 XX  
 PT Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,

PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle -  
XX  
PS Claim 1; SEQ ID No 116; 28pp; English.  
XX  
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all Cpg dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
SQ Sequence 11036 BP; 3112 A; 312 C; 2270 G; 5335 T; 7 other;  
Query Match 7.4%; Score 37; DB 22; Length 11036;  
Best Local Similarity 60.4%; Pred. No. 5.8;  
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 23 AATATTTCAAGCTAGTCATTAATTTACTTGATTCGATTCGATTATGTTTAC 82  
DB 4240 AATTTTGAGAAAGTTTGTGATTTATTTATTAATTAAGTATTTATTTTATTTA 4299  
QY 83 ATAAATTAGATTGATTTCGTGTTCTCCAAAGTCATCCAAAT 123  
DB 4300 ATTAATAATTTTCGATTTTATTTATTTAAGTGTTTAAAT 4340  
RESULT 4  
ABK28264  
ID ABK28264 standard; DNA; 11036 BP.  
XX  
AC ABK28264;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE DNA transcription associated complementary genomic DNA #69.  
XX  
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
KW immunological disorder; Werner syndrome; developmental disorder;  
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
KW neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;  
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
KW angiodysplasia; congenital heart disease; HDK syndrome; gene therapy;  
KW polyglutamine disorder; solid tumour.  
XX  
OS Unidentified.  
XX  
PN WO200192565-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 06-APR-2001; 2001WO-EP03973.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Pienbrock C, Berlin K;  
XX

DR WPI; 2002-090046/12.  
XX  
PT New nucleic acids or oligomers, useful for diagnosing or treating  
PT diseases associated with DNA transcription, e.g. immunological  
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
PT tumours or cancer  
XX  
PS Claim 1; SEQ ID No 138; 32pp; English.  
XX  
CC The invention relates to a nucleic acid, which comprises a segment of the  
CC chemically pretreated DNA of genes associated with DNA transcription from  
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
CC to the chemically pretreated DNA of genes associated with DNA  
CC transcription. The set of oligomer probes are useful for detecting the  
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
CC diagnosing or treating diseases associated with DNA transcription  
CC (particularly with the methylation status), e.g. adenosine deaminase  
CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
CC haematological disorders, immunological disorders, Werner syndrome,  
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
CC neurological disorders, neurodegenerative disorders, Wardenburg  
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart  
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
CC associated genomic DNA molecules of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 11036 BP; 3112 A; 312 C; 2270 G; 5335 T; 7 other;  
Query Match 7.4%; Score 37; DB 24; Length 11036;  
Best Local Similarity 60.4%; Pred. No. 5.8;  
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 23 AATATTTCAAGCTAGTCATTAATTTACTTGATTCGATTCGATTATGTTTAC 82  
DB 4240 AATTTTGAGAAAGTTTGTGATTTATTTATTAATTAAGTATTTATTTTATTTA 4299  
QY 83 ATAAATTAGATTGATTTCGTGTTCTCCAAAGTCATCCAAAT 123  
DB 4300 ATTAATAATTTTCGATTTTATTTATTTAAGTGTTTAAAT 4340  
RESULT 5  
ABK40057/C  
ID ABK40057 standard; DNA; 7441 BP.  
XX  
AC ABK40057;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Human chemically pretreated gene sequence #70 strand 1.  
XX  
KW Human; ds; bisulphite treatment; Cpg; DNA methylation; cancer; tumour;  
KW cytosinatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DYPD; EPHX2; OCLN; TXNRD1;  
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
XX  
OS Homo sapiens.  
XX  
PN WO200202806-A2.  
XX  
PD 10-JAN-2002.  
XX  
PF 29-JUN-2001; 2001WO-EP07470.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX





Db 4468 ATCCAT 4463

RESULT 7  
ABN71527/C  
ID ABN71527 standard; DNA: 2155561 BP.  
XX  
XX ABN71527;  
AC  
XX 02-JUL-2002 (first entry)  
DT  
XX  
DE Streptococcus polynucleotide SEQ ID NO 10967.  
XX  
XX Streptococcus; GAS; GAS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
OS Streptococcus sp.  
XX  
XX W0200234771-A2.  
XX  
XX 02-MAR-2002.  
PD  
XX 29-OCT-2001; 2001MO-GB04789.  
XX  
XX 27-OCT-2000; 2000GB-0026333.  
XX  
XX 24-NOV-2000; 2000GB-0028727.  
PR  
XX 07-MAR-2001; 2001GB-0005640.  
PR  
XX (CHIR-) CHIRON SPA.  
PA  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI  
XX Tectelin H;  
XX  
XX WPI: 2002-352536/38.  
DR  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
PT  
XX  
XX Claim 8; Page 4196-4488; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and  
CC antibodies that bind (1) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (1) are used to detect Streptococcus in a  
CC biological sample. (1) is used to determine whether a compound binds to  
CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (1) may be used to recombinantly produce (1) and may be  
CC used in gene therapy. Antibodies to (1) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
XX  
XX Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 7.3%; Score 36.4; DB 24; Length 2155561;  
Best Local Similarity 55.6%; Pred. No. 49;  
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

11 TAAAGACACCAATATTTCAGCAAGCAATTAATTTACTTGAATTCATTCGCA 70  
Db 619320 TAAATTTAAATTAACCTTTAATAAATTAATTAATTAATTCACACTTAATTAATTAATCA 619261  
QY 71 TTATGTGTACATAAATTAATTTGATTTGCTGCTCCAAAGTCATTCATTAATTCAT 130  
||||| ||||| | | | | ||||| | | | | ||| | | |

Db 619260 TTATTTATGTTAAATAGTAACAAGATTAATTCATCTTTGTCAACATTTTCAATATGACAT 619201  
QY 131 CCATTT 136  
Db 619200 CTGCTT 619195

RESULT 8  
AAT83941/C  
ID AAT83941 standard; DNA: 924 BP.  
XX  
XX AAT83941;  
AC  
XX 26-AUG-1998 (first entry)  
DT  
XX  
DE DNA encoding a molybdate-binding periplasmic protein precursor.  
XX  
XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome; ss.  
XX  
XX Staphylococcus aureus.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 119..442  
FT /\*\*tag= a  
XX  
XX W09730070-A1.  
XX  
XX 21-AUG-1997.  
PD  
XX 19-FEB-1997; 97WO-US02318.  
XX  
XX 20-FEB-1996; 96US-0011888.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
XX  
XX WPI: 1997-424969/39.  
DR P-PSDB: AAM27982.  
XX  
XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
PT to isolate antimicrobial compounds, and in vaccines against S.  
PT aureus infection  
PT  
XX  
XX Claim 9; Page 764-765; 989pp; English.

The present sequence encodes a Staphylococcus aureus protein, that,  
CC based on homology with a E. coli protein, is believed to be a  
CC molybdate-binding periplasmic protein precursor. The present sequence  
CC was obtained from a library of clones of S. aureus WCUH 29 in E. coli.  
CC The DNA sequence can be used in the construction of ribozymes and  
CC antisense sequences to control the expression of Staphylococcal genes.  
CC The DNA sequence is also useful as a source of regulatory elements for  
CC the control of bacterial gene expression. The encoded protein may be used  
CC to produce vaccines to enable a host to produce specific antibodies  
CC with antibacterial action. These vaccines and antibodies would protect  
CC a host against invasion by S. aureus, and conditions relating to  
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
CC skin syndrome, and toxic shock syndrome.  
XX  
XX Sequence 924 BP; 371 A; 108 C; 160 G; 250 T; 35 other;

Query Match 7.2%; Score 36.2; DB 18; Length 924;  
Best Local Similarity 51.7%; Pred. No. 4.1;  
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

25 TATTTCAAGACGTAGCATTAATTAATTTACTTGAATTCATTCGCAATTAATTTGTTATACAT 84  
Db 536 TAATTTGAATCTTTATACACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 477  
||||| ||||| | | | | ||||| | | | | ||||| | | | |

OY 85 AATATGATTGATTGCTGCTCCAGTCATCATTAATTCATTCATTGGACCA 144  
DB 476 CAATGCGCTNNATTCCTGCTCTNNNAATGACATCTACATCTTAGNANTNGNAGNCA 417

## RESULT 9

AAV53383/c  
ID AAV53383 standard; DNA; 924 BP.

AAV53383;

30-OCT-1998 (first entry)

DE DNA encoding a Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein: immune response induction; eye infection;  
KM antibody production; T-cell immune response; gastrointestinal infection;  
KM respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KM central nervous system; kidney infection; urinary tract infection;  
KM antimicrobial compound identification; broad spectrum antibiotic;  
KM therapy; ss.

XX Staphylococcus aureus.

OS EP841394-A2.

PD 13-MAY-1998.

PF 24-SEP-1997; 97EP-0307485.

XX 24-SEP-1996; 96US-0027032.

PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.

PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC;  
PI Lonetto MA, Nicholas RO, Pratt JM, Relchard RW, Rosenberg M;  
PI Ward JM;

DR WPI: 1998-252940/23.  
P-PSDB: AAW77585.

PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system

PS Claim 1; Page 80-81; 390pp; English.

XX This sequence encodes a Staphylococcus aureus protein of unknown.  
CC function, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.

XX Sequence 924 BP; 371 A; 108 C; 160 G; 250 T; 35 other;

Query Match 7 2%; Score 36.2; DB 19; Length 924;  
Best Local Similarity 51.7%; Pred. No. 4.1;  
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 25 TATTTCAAGCTAAGCTAATTAATTTACTGATTCATTCATTTGTTGTTACAT 84  
DB 536 TATTTGAATCTTTTATCAACATTAATCTAGACTATATNNCGCATATTATATATAT 477

OY 85 AATATGATTGATTGCTGCTCCAGTCATCATTAATTCATTCATTGGACCA 144  
DB 476 CAATGCGCTNNATTCCTGCTCTNNNAATGACATCTACATCTTAGNANTNGNAGNCA 417

## RESULT 10

AAS36807/c  
ID AAS36807 standard; DNA; 993 BP.

AAS36807;

17-DEC-2001 (first entry)

DE Human cardiovascular system antigen genomic DNA SEQ ID No 2307.

XX Cardiovascular system antigen: human; mouse; rabbit; goat; horse; cat;  
KM chicken; sheep; immunosuppressive; antiarthritis; vasotropic; dog;  
KM antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KM cerebroprotective; nocitropic; antibacterial; virucide; fungicide; cancer;  
KM ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KM cerebrovascular disorder; nervous system disorder; bacterial infection;  
KM fungal infection; viral infection; ocular disorder; endocrine disorder;  
KM gastrointestinal disorder; renal disorder; respiratory disorder;  
KM wound healing; skin aging; organ transplantation; tissue regeneration;  
KM anti-infertility.

XX Homo sapiens.

OS WO200155321-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01340.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180528.  
PR 24-FEB-2000; 2000US-0184564.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205315.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.

PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
XX	WPI; 2001-451930/48.	
DR		
XX		
PT	New cardiovascular system related polynucleotides and polypeptides,	
PT	useful for diagnosing, treating and/or preventing disorders of the	
PT	cardiovascular system -	
PS		
Claim 1; SEQ ID NO 2307; 674pp; English.		
XX		
XX	Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode	
CC	the cardiovascular system antigen polypeptides of the invention.	
CC	Cardiovascular system antigens and their associated polynucleotides are	
CC	useful in the diagnosis, treatment and prevention of various types of	
CC	disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,	
CC	chickens or sheep. A pathological condition can be determined by	
CC	detecting the presence or absence of a mutation in a cardiovascular	
CC	system antigen polynucleotide. The treatable disorders include autoimmune	
CC	diseases such as rheumatoid arthritis, hyperproliferative disorders such	
CC	as neoplasms of the breast or liver, cardiovascular disorders such as	
CC	cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,	
CC	nervous system disorders such as Alzheimer's disease, infections caused	
CC	by bacteria, viruses and fungi, ocular disorders such as corneal	
CC	infection, endocrine disorders such as premature labour and infertility,	
CC	gastrointestinal disorders such as Crohn's disease, renal disorders such	
CC	as glomerulonephritis and respiratory disorders, such as asthma and	
CC	psoriasis. The polypeptides can also be used to aid wound healing, to	
CC	prevent skin aging due to sunburn, to maintain organs before	
CC	transplantation, to regenerate tissues and in chemotaxis.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pcl_sequences.	
XX		
Query Match	7.2%; Score 36.2; DB 22; Length 993;	
Best Local Similarity	53.1%; Pred. NO. 4.2;	
Matches	77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;	
OY	23 AAATATTTCACAGCTAGCATTAATTTTACTTGGAATTCAAATTCGCATTATGTTAC 82	
Dd	677 AAAAATGTATAAAAAAATCAACAATACTTTTAGCTCATTATAGAATAATAGTACAGCTAGAG 618	
OY	83 ATAAATTAGATTGGATTTCGTGTTCTCCAAGTCCATCCAAATTAATCAATCATTGGCACG 142	
Dd	617 AGAAGATATTATATAGCATGTGCGCTTAAAGTCCATCTCAATGTTTTCATTAATAAG 558	

QY 143 CAATCTCTATGTGTCATCAT 167  
Db 557 TACATCTGATTACATATACAAACAT 533

RESULT 11  
AAS36808/c  
ID AAS36808 standard; DNA: 993 BP.  
XX  
AC AAS36808;  
XX  
DT 17-DEC-2001 (first entry)  
XX

DE Human cardiovascular system antigen genomic DNA SEQ ID No 2308.  
XX

KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW antilethemic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; noctropic; antibacterial; vitucide; fungicide; cancer;  
KW ophthalmological; vulnetary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-infertility.

XX Homo sapiens.  
XX

PN WO200155321-A2.  
XX

PD 02-AUG-2001.  
XX

PF 17-JAN-2001; 2001WO-US01340.  
XX

PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0228209.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SW;  
XX  
DR WPI; 2001-451930/48.

PT New cardiovascular system related polynucleotides and polypeptides,  
PT useful for diagnosing, treating and/or preventing disorders of the  
PT cardiovascular system -  
XX

PS Claim 1; SEQ ID No 2308; 674pp; English.

XX  
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode  
CC the cardiovascular system antigen polypeptides of the invention.  
CC Cardiovascular system antigens and their associated polynucleotides are  
CC useful in the diagnosis, treatment and prevention of various types of  
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
CC chickens or sheep. A pathological condition can be determined by  
CC detecting the presence or absence of a mutation in a cardiovascular  
CC system antigen polynucleotide. The treatable disorders include autoimmune  
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such  
CC as neoplasms of the breast or liver, cardiovascular disorders such as  
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,  
CC nervous system disorders such as Alzheimer's disease, infections caused  
CC by bacteria, viruses and fungi, ocular disorders such as corneal  
CC infection, endocrine disorders such as premature labour and infertility,  
CC gastrointestinal disorders such as Crohn's disease, renal disorders such  
CC as glomerulonephritis and respiratory disorders such as asthma and  
CC pleurisy. The polypeptides can also be used to aid wound healing, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, to regenerate tissues and in chemotaxis.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

	Query Match	Similarity	Score	DB	Length
Best Local	7.2%	53.1%	36.2	22	993
Matches	Conservative	0	Mismatches	68	Indels
					Gaps
					0
Qy	23	AAATTTTCAAGCGAAGTCATAAATTAATTTGATTCATATTCGCATTTAGCTTAC	82		
Db	677	AAAATATGAAAAAAAATTCACATACATTATGTCTAATGAATATATAGTACAGGTAGAG	618		
Qy	83	ATAAATTGATTTGATTTTCGTCTCTCCAGATTCATTCATATTCATCATTTTGGCAGC	142		

Dd		617	AGAAAGATTTTATTCAGGATGGCTCTTAAGTCATCATTGTCTTTTGCAATATAATG	558
Oy		143	CAAATCCTTATTGGTCATACAT	167
		I I I I	I I I I	
Db		557	TACATCGATTACATATACAACAT	533
 RESULT 12 ABL1650 ID ABL16650 standard; DNA: 26636 BP.  XX ABL1650; AC XX DE 26-MAR-2002 (first entry) XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1423. XX DE Drosophila developmental biology; cell signalling; insecticide; KW pharmaceutical; gene; ds. XX OS Drosophila melanogaster. XX PN W0200171042-A2. FN PD 27-SEP-2001. PP PP 23-MAR-2001; 2001WO-US09231. XX PR 23-MAR-2000; 2000US-191637P. PR 11-JUL-2000; 2000US-0614150. XX PA (PEKE ) PE CORP NY. XX PI Venter JC, Adams M, Li PWD, Myers EW, DR WIJ; 2001-656860/75.				

PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from <i>Drosophila</i> and for elucidating cell signalling and cell-cell	
PT	interactions -	
XX		
PS	Claim 1; SEQ ID NO 1423; 21np + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from <i>Drosophila</i> . The invention is	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (AB116176-AB120511), expressed DNA	
CC	sequences (AB101840-AB116175) and the encoded proteins	
CC	(AB157737-AB172072).	
CC	The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at <a href="http://wipo.int/pub/published_pct_sequences">ftp://wipo.int/pub/published_pct_sequences</a> .	
XX		
SO	Sequence 26636 BP; 7621 A; 5571 C; 5599 G; 7845 T; 0 other;	
Query Match	7.2%; Score 36; DB 23; Length 26636;	
Best Local Similarity	53.6%; Pred. NO. 15;	
Matches	75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;	
Oy	65 TTCCATATATGTTGCATCAATTAATTACATTTGATTCGTCGTCGAAGTCATCAATT 124	
Dbb	22682 TTCCCATTTCAAGTCTCGACATTTTTCATCAATTCATACGACATCTTCTTCGCTGCT 2274	
Oy	125 ATCCATCAATTTGGCAGCCAAATCCTTTATTTGGTCATACATCAATGAGTATCTTCTCA 184	
Dbb	22742 CTAATATGCTTTGGCCCCCTAAATCTCATCTACTATATATGTACATATGTATGTATGCA 2280	
Oy	185 CAAGACACAGTTAGAAGCA 204	
Dbb	22802 CACATTAATAGTTCTATATCA 22821	

RESULT 13  
AAK9724/C  
ID AAK9724 standard; DNA: 3030 BP.  
XX  
AC AAK9724;  
XX  
DT 11-JUL-2002 (first entry)  
XX  
DE DNA of the BpmI endonuclease gene (BpmIRM).  
XX  
KW BpmI restriction endonuclease; Bacillus pumilus; recombinant expression;  
KM cleaving DNA; cloning; gene characterisation; methylase; BpmIRM;  
KM BpmI endonuclease gene; gene; ds.  
XX  
OS Bacillus pumilus.  
XX  
FH Key location/Qualifiers  
FT CDS 1..3030  
FT /tag- a  
FT /product- "Protein of the BpmI endonuclease gene  
FT (BpmIRM)"  
XX  
PN EPI199365-A2.  
XX  
PD 24-APR-2002.  
XX  
PE 10-OCT-2001; 2001EP-0203826.  
XX  
PR 20-OCT-2000; 2000US-0693146.  
XX  
PA (NEW ) NEW ENGLAND BIOLABS INC.  
XX  
PI Xu S, Xiao J, Zhu Z;  
XX  
XX WPI: 2002-354206/39.  
DR P-PSDB: AAO20949.  
XX  
XX New isolated DNA encoding BpmI restriction endonuclease, useful for  
PT cleaving DNA and for the associated methylase  
XX  
PS Disclosure; Fig 3; 33pp; English.  
XX  
CC The invention relates to an isolated DNA segment encoding the BpmI  
CC restriction endonuclease and isolated from Bacillus pumilus, New England  
CC Biolabs Collection No. 711. The isolated DNA is used for recombinant  
CC expression of BpmI, and is useful as a reagent for cleaving DNA, and for  
CC cloning or gene characterisation. The associated methylase, BpmIM, can  
CC be used similarly. This polynucleotide sequence represents the DNA of the  
CC BpmI endonuclease gene (BpmIRM) of the invention.  
XX  
SQ Sequence 3030 BP; 1137 A; 405 C; 572 G; 916 T; 0 other;  
Query Match 7.1%; Score 35.8; DB 24; Length 3030;  
Best Local Similarity 52.3%; Pred. No. 8;  
Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 4 TATGACTAAAGCACCACATATTTCCAAAGCTAAGCATTAATATTACTGCAATCA 63  
DB 1339 TAAATATCAACGCAACAATAATATTCGCAAGATATTTCTTTGCTTAAGTCAAA 1280  
QY 64 ATTGCATTTATGTTTACATTAATTTAGATTGATTTGGTGTTCACAGTCCATCCAAAT 123  
DB 1279 TTACCCCTTCTTCAGTTTCAATATCTGAATCATGATGTCTCTCTTATATCT 1220  
QY 124 TATGATCCATTTGGGACGCCAAATCCCTTA 154  
DB 1219 TTTCATATTACTTCTCTACTAGAAAGTCATA 1189

XX  
AC AAK9724/C  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE Human cardiovascular system antigen cdNA polynucleotide SEQ ID No 547.  
XX  
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; nocrotropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-infertility.  
XX  
OS Homo sapiens.  
XX  
PN WO200155321-A2.  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01340.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0220954.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227189.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 03-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.





Job time : 169.778 secs

```
ID    AAH54882 standard; DNA: 3030 BP.
XX
AC    .AAH54882;
XX
DT    03-SEP-2001 (first entry)
XX
DE    S. epidermidis genomic polynucleotide sequence SEQ ID NO:4246.
XX
KW    Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX    vaccination; endocarditis; ds.
XX
OS    Staphylococcus epidermidis.
XX
PN    W0200134809-A2.
XX
PD    17-MAY-2001.
XX
PF    09-NOV-2000; 2000WO-US30782.
XX
PR    09-NOV-1999; 99US-0164258.
XX
PA    (GLAX ) GLAXO GROUP LTD.
XX
PI    Kimmerly WJ;
XX
DR    WPI: 2001-316495/33.
XX
PT    Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX    useful for vaccinating against infections, e.g. endocarditis -
XX
PS    Claim 8; Page 1968-1969; 2188pp; English.
XX
CC    AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC    (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC    (I) and (II) can have antibacterial activity and therefore can be used
CC    in vaccination. The nucleic acids (I) may be used to produce the
CC    S. epidermidis polypeptides (II) via the production of vectors
CC    containing them which are used to produce hosts cells which express the
CC    polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC    used to vaccinate subjects and to raise antibodies against the bacteria.
CC    The polypeptides may also be used to assay for other inhibitors of their
CC    activity and therefore identify compounds that may be used for the
CC    treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC    AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC    polynucleotide sequences from the present invention. AAH55091 to
CC    AAH55098 represent oligonucleotide sequences and primers which are used
CC    in the exemplification of the present invention.
CC    N.B. The present invention specifically claims all the polynucleotide
CC    sequences given in the sequence listing of the present specification,
CC    however the sequence listing only goes up to SEQ ID NO:4454 so even
CC    though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC    no sequences are present for SEQ ID NO:4455 to 4464.
XX
SO    Sequence 3030 BP: 1135 A; 419 C; 558 G; 918 T; 0 other;

Query Match      7.1%; Score 35.4; DB 22; Length 3030;
Best Local Similarity 53.2%; Pred. No. 10;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY    15 AGCACCAATATTTCAGCTAGTCATTAATATTACTTGATGATCAATTCGATTTAT 74
DB    1441 AACTGCACAAATTTTAAAGATGATCTCATCTAGTACGATTCGCAAAATPAG 1500
OY    75 TGTGTACATAAATTAGATTGTTGCTTCAGATCATCATTCATTCATCCAT 134
DB    1501 TAGGTAAATATAAAGAGAGTACTGATGATCATCATGACTATTAATTTACTGTAAT 1560
OY    135 TTGGCAGCCAAATCCTCTAT 155
DB    1561 TTCTTATCTAATAAGTTAAT 1581
```

Search completed: July 4, 2003, 15:25:04

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 15:08:57 ; Search time 1508.67 Seconds

(without alignments)  
9664.474 Million cell updates/sec

US-10-083-853b-2\_COPY\_18900\_19400

Sequence: 1 ttctatgaactaaagcacc.....gctccctgttcttttt 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

GenBank: 1: gb\_ba: 2: gb\_htg: 3: gb\_in: 4: gb\_ov: 5: gb\_ov: 6: gb\_ph: 7: gb\_ph: 8: gb\_pl: 9: gb\_pl: 10: gb\_ro: 11: gb\_ro: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_mu: 21: em\_or: 22: em\_ov: 23: em\_ov: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_ro: 28: em\_un: 29: em\_vl: 30: em\_vl: 31: em\_vl: 32: em\_vl: 33: em\_vl: 34: em\_vl: 35: em\_vl: 36: em\_vl: 37: em\_vl: 38: em\_vl: 39: em\_vl: 40: em\_vl: 41: em\_vl:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	180648	9 AP003043	AP003043 Homo sapi
2	501	100.0	182403	2 AC024052	AC024052 Homo sapi
3	65	13.0	160345	2 AC067870	AC067870 Homo sapi
4	43	8.6	176784	2 AC016632	AC016632 Homo sapi
5	43	8.6	182126	2 AC016639	AC016639 Homo sapi
6	41.4	8.3	193087	2 AC010779	AC010779 Homo sapi
7	40.4	8.1	149628	2 AC080186	AC080186 Homo sapi
8	40.4	8.1	172868	2 AC019140	AC019140 Homo sapi
9	40.4	8.1	176951	2 AL845340	AL845340 Mus muscu
10	40.2	8.0	227054	9 AF172277	AF172277 Homo sapi
11	40.2	8.0	227054	2 AF172277	AF172277 Homo sapi
12	40	8.0	55470	2 AC116989	AC116989 Dictyoste
13	40	8.0	150950	2 AC119464	AC119464 Rattus no
14	39.8	7.9	103138	2 AC112117	AC112117 Rattus no
15	39.8	7.9	139999	8 AC018727	AC018727 Oryza sat
16	39.8	7.9	163166	9 AL356632	AL356632 Human DNA
17	39.8	7.9	171014	10 AL671915	AL671915 Mouse DNA
18	39.8	7.9	188993	2 AC109656	AC109656 Rattus no
19	39.6	7.9	3457	3 AF310886	AF310886 Dictyoste
20	39.6	7.9	4197	3 AF163835	AF163835 Dictyoste
21	39.2	7.8	107341	10 AL805932	AL805932 Mouse DNA
22	39.2	7.8	138397	2 AC095231	AC095231 Rattus no
23	39.2	7.8	148391	2 AC034174	AC034174 Homo sapi
24	39.2	7.8	159012	2 AC023145	AC023145 Homo sapi
25	39.2	7.8	159972	2 AC107810	AC107810 Mus muscu
26	39.2	7.8	160250	9 AC087173	AC087173 Homo sapi
27	39.2	7.8	160454	9 AC087366	AC087366 Homo sapi
28	39.2	7.8	163717	2 AC127281	AC127281 Mus muscu
29	39.2	7.8	192404	9 CENS1DWO	AL136296 Human chr
30	39.2	7.8	194985	10 AC002406	AC002406 Human chr
31	39	7.8	61513	9 AL353610	AL353610 Human DNA
32	39	7.8	160324	9 AC079117	AC079117 Homo sapi
33	38.8	7.7	43834	3 AF045643	AF045643 Caenorhab
34	38.8	7.7	66679	2 AC118587	AC118587 Mus muscu
35	38.8	7.7	161495	2 AC120893	AC120893 Rattus no
36	38.8	7.7	171668	2 AC112975	AC112975 Mus muscu
37	38.8	7.7	187023	2 AC006882	AC006882 Caenorhab
38	38.6	7.7	67247	2 AC121357	AC121357 Mus muscu
39	38.4	7.7	104333	9 AL133295	AL133295 Human DNA
40	38.4	7.7	138159	2 AP002742	AP002742 Homo sapi
41	38.4	7.7	164301	2 AC018671	AC018671 Homo sapi
42	38.4	7.7	164513	9 AC093262	AC093262 Homo sapi
43	38.4	7.7	164996	2 AC087372	AC087372 Homo sapi
44	38.4	7.7	197665	9 AC108456	AC108456 Homo sapi
45	38.2	7.6	27908	3 CETO4H1	278200 Caenorhabd1

# ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AP003043	AP003043	Homo sapiens genomic DNA, chromosome 11q clone:RP11-617B3, complete	AP003043	AP003043.2	GI:15637094	Homo sapiens DNA, clone:RP11-617B3.	Homo sapiens	1	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA  
JOURNAL Published Only in Database (2000)  
REFERENCE 2 (bases 1 to 180648)  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tokoki,Y., Matsuabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT On Sep 17, 2001 this sequence version replaced gi:11862945.  
FEATURES  
source  
1.180648  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11q"  
/clone="RP11-617B3"  
BASE COUNT 56234 a 35375 c 33753 g 55286 t  
ORIGIN  
Query Match 100.0%; Score 501; DB 9; Length 180648;  
Best Local Similarity 100.0%; Pred. NO.4.5e-124;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTCTATGAACATAAGACACCAACATATTTTCAAGCTAAGTCAATAATATTACTTGAAT 60  
DB 58769 TTCTATGAACATAAGACACCAACATATTTTCAAGCTAAGTCAATAATATTACTTGAAT 58828  
QY 61 TCATTCGCATTATTTGTGTACATTAATTAATGATTTGATTTGCTTCTCCAGTCAATCC 120  
DB 58829 TCATTCGCATTATTTGTGTACATTAATTAATGATTTGATTTGCTTCTCCAGTCAATCC 58888  
QY 121 AATTATCCATTCATTTGGGACCAATTCCTTATTTGGTGCATACATACATGATTAATCTT 180  
DB 58889 AATTATCCATTCATTTGGGACCAATTCCTTATTTGGTGCATACATACATGATTAATCTT 58948  
QY 181 CTCACAGAAGACAGTATGAAGCAAGTCTTTTGTCTGCACAGAACCCACCCCTTTCC 240  
DB 58949 CTCACAGAAGACAGTATGAAGCAAGTCTTTTGTCTGCACAGAACCCACCCCTTTCC 59008  
QY 241 CAATGACGCCCTGTGTGACGACGACCAACAGAAAGAAATCCATGATTTAGGACG 300  
DB 59009 CAATGACGCCCTGTGTGACGACGACCAACAGAAAGAAATCCATGATTTAGGACG 59068  
QY 301 TTATCACCATTGTCATTTAACTCTAGCCCTGTAAATGTCTTAGGCTTAACTGAAC 360  
DB 59069 TTATCACCATTGTCATTTAACTCTAGCCCTGTAAATGTCTTAGGCTTAACTGAAC 59128  
QY 361 TCGTTAAGGAGAAGAAAAAAGCCCTTTGGGACGAGAGTCAAAAGTCAAGACAGT 420  
DB 59129 TCGTTAAGGAGAAGAAAAAAGCCCTTTGGGACGAGAGTCAAAAGTCAAGACAGT 59188  
QY 421 GTCAGTAGAATGATTAAGAAAAATTTGAGAGAAAGTGGACCATAGTCCATTAAC 480  
DB 59189 GTCAGTAGAATGATTAAGAAAAATTTGAGAGAAAGTGGACCATAGTCCATTAAC 59248  
QY 481 CGCATCCCTGTTTGTCTTTT 501  
DB 59249 CGCATCCCTGTTTGTCTTTT 59269

RESULT 2  
AC024052 182403 bp DNA linear HTG 17-AUG-2000-  
LOCUS Homo sapiens chromosome 11 clone RP11-617B3, WORKING DRAFT  
DEFINITION  
SEQUENCE, 7 unordered pieces.  
ACCESSION AC024052  
VERSION AC024052.3 GI:9838295  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Tokoki,Y., Matsuabe,H. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT On Sep 17, 2001 this sequence version replaced gi:11862945.  
FEATURES  
source  
1.182403  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-617B3"  
misc\_feature  
1.1666  
/note="assembly\_name:Contig2"  
1767.9133  
/note="assembly\_name:Contig3  
vector\_side:left"  
9234.20663  
/note="assembly\_name:Contig4"  
20764.35927  
/note="assembly\_name:Contig5"  
36028.61610  
/note="assembly\_name:Contig6"

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will be preserved. \*\*\*\*\*  
1 1666: contig of 1666 bp in length  
\* 1667 1766: gap of unknown length  
\* 1767 9133: contig of 7367 bp in length  
\* 9134 9233: gap of unknown length  
\* 9234 20663: contig of 11430 bp in length  
\* 20664 20763: gap of unknown length  
\* 20764 35927: contig of 15164 bp in length  
\* 35928 36027: gap of unknown length  
\* 36028 61610: contig of 25583 bp in length  
\* 61611 61710: gap of unknown length  
\* 61711 96167: contig of 34457 bp in length  
\* 96168 96267: gap of unknown length  
\* 96268 182403: contig of 86136 bp in length.  
\*\*\*\*\*  
Location/Qualifiers  
1.182403  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-617B3"  
1.1666  
/note="assembly\_name:Contig2"  
1767.9133  
/note="assembly\_name:Contig3  
vector\_side:left"  
9234.20663  
/note="assembly\_name:Contig4"  
20764.35927  
/note="assembly\_name:Contig5"  
36028.61610  
/note="assembly\_name:Contig6"

```

misc_feature      61711..96167
                  /note="assembly_name:Contig7"
misc_feature      96268..182403
                  /note="assembly_name:Contig8
                  vector_side:left"
BASE COUNT      55923 a 35190 c 34519 g 56169 t 602 others
ORIGIN
Query Match      100.0%; Score 501; DB 2; Length 182403;
Best Local Similarity 100.0%; Pred. No. 4,5e-124;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTCTAGACATTAAGACACCAATATTTTCAAGCTATGATTAATTTACTTGAT 60
DB 155014 TTCTATGACATTAAGACACCAATATTTTCAAGCTATGATTAATTTACTTGAT 155073
OY 61 TCAATTCGATTAATTTGTGTACATTAATTTAGATTTGTTGTTTCAAGTCCATCC 120
DB 155074 TCAATTCGATTAATTTGTGTACATTAATTTAGATTTGTTGTTTCAAGTCCATCC 155133
OY 121 AATTATCCATCTTTGGACGCAATTCCTTATTTGTGATATCATGAGATCTT 180
DB 155134 AATTATCCATCTTTGGACGCAATTCCTTATTTGTGATATCATGAGATCTT 155193
OY 181 CTCACAGAAGACAGATTAGAGCAAGTCTTTTGTCTGACAGAAACCACCCCTTCC 240
DB 155194 CTCACAGAAGACAGATTAGAGCAAGTCTTTTGTCTGACAGAAACCACCCCTTCC 155253
OY 241 CAATGACAGCCCTTGTGTGACACGACGACACAGAAAGAAATCCATGATTTAGAGCAG 300
DB 155254 CAATGACAGCCCTTGTGTGACACGACGACACAGAAAGAAATCCATGATTTAGAGCAG 155313
OY 301 TTATCACCATGTTATCTTAACTTACCTGTAATGTCTTAGGCTTAACGTGAAC 360
DB 155314 TTATCACCATGTTATCTTAACTTACCTGTAATGTCTTAGGCTTAACGTGAAC 155373
OY 361 TGTCTTAAGAGAGAGAGAAAGAAAGCCTTTGGAGAGAGTCAAAAGTCAAGAGACGT 420
DB 155374 TGTCTTAAGAGAGAGAGAAAGAAAGCCTTTGGAGAGAGTCAAAAGTCAAGAGACGT 155433
OY 421 GTCACTAGATGATCTTAAGAAATTTGACAGAAACAGTGTGACCATGACTCCACTAAC 480
DB 155434 GTCACTAGATGATCTTAAGAAATTTGACAGAAACAGTGTGACCATGACTCCACTAAC 155493
OY 481 CGCATCCCTGTTTGTCTTTT 501
DB 155494 CGCATCCCTGTTTGTCTTTT 155514

RESULT 3
AC067870 160345 bp DNA linear HTG 23-JUN-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-475J2 map 11, WORKING DRAFT
SEQUENCE 18 unordered pieces.
AC067870
AC067870.3 GI:8671974
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 160345)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 11, clone RP11-475J2
Unpublished
2 (bases 1 to 160345)
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beka,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Deatrelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

```

```

Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vaasiliu,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 23, 2000 this sequence version replaced gi:8099852.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L8810
Center clone name: 475_J-2
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151812 bases at least Q40
Consensus quality: 156075 bases at least Q30
Insert size: 157717 bases at least Q20
Insert size: 158645; sum-of-coverage
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 946: contig of 946 bp in length
* 947 1046: gap of 100 bp
* 1047 2803: contig of 1757 bp in length
* 2804 2903: gap of 100 bp
* 2904 4474: contig of 1571 bp in length
* 4475 4574: gap of 100 bp
* 4575 8112: contig of 3538 bp in length
* 8113 8212: gap of 100 bp
* 8213 11442: contig of 3330 bp in length
* 11443 11542: gap of 100 bp
* 11543 14027: contig of 2485 bp in length
* 14028 14127: gap of 100 bp
* 14128 17482: contig of 3355 bp in length
* 17483 17582: gap of 100 bp
* 17583 20007: contig of 2425 bp in length
* 20008 20107: gap of 100 bp
* 20108 26548: contig of 6441 bp in length
* 26549 26648: gap of 100 bp
* 26649 34423: contig of 7775 bp in length
* 34424 34523: gap of 100 bp
* 34524 41559: contig of 7036 bp in length
* 41560 41659: gap of 100 bp
* 41660 50886: contig of 9227 bp in length

```



OY	2	TCTTTGACCTAAAGACACACATATTTTCAAGCTAGTCATTAATTTTACTTGAA	161
Db	123123	TCTAAATTTTAAACCAACATRTATTTTCATCTCAATAACATRTATTTACTTAGAT	123005
OY	62	CAATTCGCATTTATGTGTTACATTAATTTAGATTTGAA	100
Db	123063	CATTTAGTATCATGATTTACAAATTCCTTGCAATTTGAT	123025
RESULT 5			
AC016639/c			
LOCUS	AC016639	182126 bp	DNA
DEFINITION	Homo sapiens chromosome 5 clone RP11-412L4, WORKING DRAFT SEQUENCE,		
	7 ordered pieces.		
ACCESSION	AC016639		
VERSION	AC016639.5 GI:7711585		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	DOE Joint Genome Institute.		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Sequencing of Human Chromosome 5		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 182126)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On May 6, 2000 this sequence version replaced gi:7710162.		





misc.feature 21414..25335  
/note="assembly-fragment"  
misc.feature 25436..32017

Query Match 8.38; Score 41.4; DB 2; Length 193087;  
Best Local Similarity 50.28; Pred. No. 1.6;  
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

OY 2 TCTATGACATAAAGCAGCACCAATATTTTCAAGCTAGTCATTAATATTACTGAT 61  
Db 133622 TATATTTCTTATAGTACGCTTACTATACAGGCTTATAGTTACTATTTTTCATATAT 133681  
OY 62 CAATTCGCAATATTTGCTTACATTAATAGTTGATTTGCTTCTCAAGTCCATCA 121  
Db 133682 TATTTTATGCAATTTTAAAGTATGTAATTTCTCTCTCTTCTTCTCAATCTGCTCA 133741  
OY 122 ATATTCATTCATTTGGCAGCCAAATCTCTTATGTCGATACATACATGATCTTC 181  
Db 133742 AGTCACCTTCCCTCACTTCACTTCTGCAATTAATATATATACATATATATAT 133801  
OY 182 TCACAAAGACAGGTTGAAGCA 204  
Db 133802 ACATACATTAATATATATATCA 133824

RESULT 7  
AC080186/c 149628 bp DNA linear HTG 18-FEB-2001  
LOCUS Homo sapiens chromosome 2 clone RP11-69M17, WORKING DRAFT SEQUENCE,  
15 unordered pieces.  
AC080186  
AC080186.2 GI:12963034  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 149628)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 149628)  
Waterston, R.H.  
Direct Submission  
Submitted (28-SEP-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Feb 18, 2001 this sequence version replaced gi:10334942.

COMMENT ----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Project Information -----  
Center project name: H.NH0069M17  
Summary Statistics -----  
Sequencing vector: p1asmid; 898;  
Chemistry: Dye-Primer ET; 89% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 140807 bases at least Q40  
Consensus quality: 143657 bases at least Q30  
Consensus quality: 144902 bases at least Q20  
Insert size: 148000; agarose-fp  
Insert size: 148228; sum-of-contigs  
Quality coverage: 4.36 in Q20 bases; sum-of-contigs  
Quality coverage: 4.39 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6327: contig of 6327 bp in length  
6328 6427: gap of unknown length  
6428 25696: contig of 19266 bp in length  
25697 25797: gap of unknown length  
25797 40438: contig of 14642 bp in length  
40438 40539: gap of unknown length  
40539 55132: contig of 14593 bp in length  
55132 55232: gap of unknown length  
55232 77739: contig of 22408 bp in length  
77739 77740: gap of unknown length  
77740 100829: contig of 23099 bp in length  
100829 100929: gap of unknown length  
100929 126647: contig of 25718 bp in length  
126647 126748: gap of unknown length  
126748 128819: contig of 2072 bp in length  
128819 128920: gap of unknown length  
128920 130922: contig of 2003 bp in length  
130922 131022: gap of unknown length  
131022 132325: contig of 1303 bp in length  
132325 132426: gap of unknown length  
132426 134035: contig of 1610 bp in length  
134035 134135: gap of unknown length  
134135 137686: contig of 3551 bp in length  
137686 137787: gap of unknown length  
137787 140534: contig of 2748 bp in length  
140534 140635: gap of unknown length  
140635 143274: contig of 2640 bp in length  
143274 143375: gap of unknown length  
143375 149628: contig of 6254 bp in length.

FEATURES  
Source  
1..149628  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-69M17"

1..6327  
/note="assembly\_name:Contig10  
clone\_end:SP6  
vector\_side:right"  
6428..25696  
/note="assembly\_name:Contig13  
clone\_end:T7  
vector\_side:left"  
25797..40438  
/note="assembly\_name:Contig11"  
40539..55131  
/note="assembly\_name:Contig12"  
55232..77639  
/note="assembly\_name:Contig14"  
77740..100829  
/note="assembly\_name:Contig15"  
100930..126647  
/note="assembly\_name:Contig16"  
126748..128819  
/note="assembly\_name:Contig12"  
128920..130922  
/note="assembly\_name:Contig3"  
131023..132325  
/note="assembly\_name:Contig4"  
132426..134035  
/note="assembly\_name:Contig5"  
134136..137686  
/note="assembly\_name:Contig6"  
137787..140534  
/note="assembly\_name:Contig7"  
140635..143274  
/note="assembly\_name:Contig8"  
143375..149628  
/note="assembly\_name:Contig9"



```

repeat_region      18589. .18690
/rpt_family="L2"
repeat_region      18691. .18989
/rpt_family="Alu"
repeat_region      18990. .21365
/rpt_family="L2"
repeat_region      21374. .21472
/rpt_family="MIR"
repeat_region      21732. .21812
/rpt_family="L2"
repeat_region      22068. .22615
/rpt_family="L2"
repeat_region      23058. .23403
/rpt_family="MIR"
repeat_region      23672. .23767
/rpt_family="L2"
repeat_region      25785. .26082
/rpt_family="Alu"
repeat_region      29162. .29579
/rpt_family="ERV1"
repeat_region      32263. .32394
/rpt_family="L2"
repeat_region      32464. .32611
/rpt_family="MIR"
repeat_region      32622. .32694
/rpt_family="Alu"
misc_feature        32767. .32977
/note="similar to
34502. .34765
/rpt_family="MaLR"
repeat_region      44600. .44668
/rpt_family="MER103"
repeat_region      44878. .45052
/rpt_family="MIR"
repeat_region      45370. .45614
/rpt_family="L2"
repeat_region      45918. .46175
/rpt_family="Alu"
misc_feature        45943. .46339
/note="similar to
46457. .46783
/note="similar to
47544. .47831
/rpt_family="Alu"
repeat_region      47832. .47936
/rpt_family="MIR"
repeat_region      48162. .48306
/rpt_family="MIR"
repeat_region      49505. .49642
/rpt_family="MIR"
repeat_region      49828. .50351
/rpt_family="ERV1"
repeat_region      50787. .51007
/rpt_family="MER1"
repeat_region      52402. .52491
/rpt_family="MER53"
repeat_region      52499. .52563
/rpt_family="MER53"
repeat_region      53821. .53906
/rpt_family="L2"
repeat_region      53940. .54166
/rpt_family="L2"
repeat_region      54152. .54522
/rpt_family="L2"
repeat_region      54669. .54979
/rpt_family="Alu"
repeat_region      56513. .56885
/rpt_family="MaLR"
repeat_region      56886. .58488
/rpt_family="MaLR"
repeat_region      58491. .58792
/rpt_family="MaLR"
repeat_region      58793. .59095

```

```

repeat_region      59096. .59166
/rpt_family="MaLR"
/rpt_family="MaLR"

Query Match
Best Local Similarity 54.8%; Pred. No. 3;
Matches 80; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 19 CCNCAATATTTGCAACCTAAGTCATTAATATTACTTGAAATGCATTCGCATATTGTG 78
Db 78630 CCAATATTTTGTGCTCACAATTAAGAGACATGAGTATTTGCTTCCTGTG 78689
QY 79 TTACATTAATTTAGATTGATTCGTCCTCAAGTCATTCATTAATTCATTCATTTGG 138
Db 78690 TCTGACTATTTTATTTACATATATGCTCCAGTGTCATTCATGTCATGATGACA 78749
QY 139 CAGCCAAATCCCTTATTTGTCATTA 164
Db 78750 GGATTTATTTATTTTCATGCGTGAATA 78775

RESULT 9
AL845340
LOCUS
DEFINITION
MUS musculus chromosome 4 clone RP23-169D3, *** SEQUENCING IN
PROGRESS ***, 18 unordered pieces.
ACCESSION
AL845340.1 GI:22205035
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1..to 176951)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: DMI69D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 164062 bases at least Q40
Consensus quality: 169446 bases at least Q30
Consensus quality: 172590 bases at least Q20
Insert size: 175251; sum-of-contrigs
Insert size: 180267; 5.0% error; agarose-fp
Quality coverage: 3.18x in Q20 bases; sum-of-contrigs Quality
coverage: 3.18x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2625: contig of 2625 bp in length
* 2626 2725: gap of 100 bp
* 2726 5168: contig of 2443 bp in length
* 5169 5268: gap of 100 bp
* 5269 13730: contig of 8482 bp in length
* 13751 13850: gap of 100 bp
* 13851 17241: contig of 3391 bp in length
* 17242 17341: gap of 100 bp

```



JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 5, 1999 this sequence version replaced gi:4204336.

## COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

----- Summary Statistics

Center project name: H\_RG062N11

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [malco:egreen@nhgri.nih.gov](mailto:malco:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

Clone CTB-62N11 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBel0BAC11

Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-15F3, 200 bp overlap; the clone sequenced to the right is CTB-36G2, 200 bp overlap. Actual start of this clone is at base position 195 of CTB-62N11; actual end is at 31845 of CTB-36G2.

## FEATURES

## Source

1..74673  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q22"  
/clone="CTB-62N11"  
/clone\_1lb="CTB-978SK-B"  
307..374  
/rpt\_family="L2"  
648..678  
/rpt\_family="AT\_rich"  
2294..2602  
/rpt\_family="Alu"  
2897..3017  
/rpt\_family="MIR"  
3254..3582  
/rpt\_family="MERL\_group"  
3663..3982  
/rpt\_family="MERL\_type"  
4015..4273  
/rpt\_family="L2"  
5163..5221  
repeat\_region

/rpt\_family="L2"  
5311..5708  
/rpt\_family="MaLR"  
5709..6007  
/rpt\_family="Alu"  
6008..6061  
/rpt\_family="MaLR"  
6180..6629  
/rpt\_family="MaLR"  
7642..7718  
/rpt\_family="MIR"  
7975..8370  
/rpt\_family="MaLR"  
9210..9251  
/rpt\_family="L2"  
9293..9770  
/rpt\_family="L2"  
9821..9936  
/rpt\_family="MERL\_type"  
9958..10234  
/rpt\_family="Alu"  
10322..10483  
/rpt\_family="L2"  
10487..10653  
/rpt\_family="MaLR"  
10962..11619  
/rpt\_family="L1"  
11742..11906  
/rpt\_family="MERL\_type"  
12278..12496  
/rpt\_family="L2"  
12497..12786  
/rpt\_family="Alu"  
12787..13190  
/rpt\_family="L2"  
13288..13822  
/rpt\_family="L2"  
13816..14069  
/rpt\_family="L2"  
14148..14257  
/rpt\_family="MIR"  
14421..14470  
/rpt\_family="G-rich"  
14898..14918  
/rpt\_family="AT\_rich"  
15485..15775  
/rpt\_family="Alu"  
16602..16809  
/db\_xref="GI:859023"  
17856..18120  
/db\_xref="GI:859034"  
complement(16263..18500)  
/db\_xref="GI:3660631"  
20930..21613  
/rpt\_family="L2"  
21614..21726  
/rpt\_family="MERL\_type"  
21727..22019  
/rpt\_family="L2"  
22031..23210  
/rpt\_family="L1"  
23325..23440  
/rpt\_family="L1"  
24059..24101  
/rpt\_family="(CA)n"  
24269..24297  
/rpt\_family="AT\_rich"  
24456..24708  
/rpt\_family="L2"  
26522..26621  
/rpt\_family="L1"  
30969..31166  
/rpt\_family="MIR"  
repeat\_region

repeat\_region 31955..31995  
/rpt\_family="(CA)n"  
repeat\_region 33505..33532  
/rpt\_family="(GGAGA)n"  
repeat\_region 34218..34363  
/rpt\_family="MIR"  
repeat\_region 35193..35504  
/rpt\_family="Alu"  
STS 35610..36017  
/db\_xref="GI:1408103"  
repeat\_region 39007..39288  
/rpt\_family="L1"  
repeat\_region 39289..39436  
/rpt\_family="MER4-group"  
repeat\_region 39440..39740  
/rpt\_family="MER4-group"  
repeat\_region 39741..39949  
/rpt\_family="L1"  
repeat\_region 40147..40439  
/rpt\_family="L2"

Query Match 8.0%; Score 40.2; DB 9; Length 74673;  
Best Local Similarity 56.4%; Pred. No. 3.4; Indels 0; Gaps 0;  
Matches 75; Conservative 0; Mismatches 58;

QY 24 AATTTTCAAGCTAAGTCAATAATTTACTGAAATTCGATTTGCTTACA 83

Db 67226 AATTTAGCAAGGTAGTCTCAATATCTAGTCTTAAAGTGAATGCTGTAACA 67285

QY 84 TAATATAGATTTTATTTGCTTCAGTCCATCCATATTCATTCATTGGACGC 143

Db 67286 CTAAATATTTATTTATTTTATTTTAAAGACCTTCATTCCTGCCCTT 67345

QY 144 AATTCCTCTATT 156

Db 67346 GTATCTTTT 67358

RESULT 11  
AF172277/c 227054 bp DNA linear PRI 08-JAN-2002  
LOCUS Homo sapiens 7q31 clones RP11-267017 and RP11-25612 including NRCAM  
DEFINITION and LAMB1R genes, complete sequence.  
ACCESSION AF172277  
VERSION AF172277  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 227054)

Gene 273 (1), 115-122 (2001)  
and conserved genomic organization for the L1 subfamily  
adhesion molecule reveals a novel alternative exon in chick and man

JOURNAL  
MEDLINE 21376132  
PUBMED 11483367  
REFERENCE 2 (bases 1 to 227054)  
AUTHORS Platzer M.  
TITLE Direct Submision  
JOURNAL Submitted (24-JUL-1999) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
FEATURES  
location/Qualifiers  
1..227054  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21"  
1..149111  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-267017"

source  
/clone="RP11-267017"

source  
/clone\_11b="RP11 human BAC 11"  
46608..227050  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-25612"  
/clone\_11b="RP11 human BAC 11"  
complement(743..1982)  
/rpt\_family="L1PAL16"  
complement(1978..2174)  
/rpt\_family="L1"  
2408..2448  
/rpt\_family="5S"  
2477..2585  
/rpt\_family="MIR"  
2911..3172  
/rpt\_family="MER58B"  
complement(4213..4343)  
/rpt\_family="L2"  
complement(4527..4741)  
/rpt\_family="L2"  
5016..5295  
/rpt\_family="AluSg"  
complement(6112..6288)  
/rpt\_family="MER5B"  
9282..9624  
/rpt\_family="AluJb"  
10105..10186  
/rpt\_family="MIR"  
10246..16379  
/rpt\_family="L1PA7"  
complement(16528..16664)  
/rpt\_family="MER46C"  
complement(17431..17725)  
/rpt\_family="AluSx"  
complement(17974..18041)  
/rpt\_family="L1MB4"  
complement(18081..18141)  
/rpt\_family="U7"  
complement(18142..18555)  
/rpt\_family="L1MB3"  
complement(19664..19776)  
/rpt\_family="L2"  
20074..20263  
/rpt\_family="MIR"  
complement(20612..20715)  
/rpt\_family="U6"  
21848..22299  
/rpt\_family="L1FR40b"  
complement(22646..23050)  
/note="WI-13458, G24693"  
23454..23567  
/rpt\_family="MIR"  
24896..25185  
/rpt\_family="AluSx"  
26619..26740  
/rpt\_family="MIR"  
complement(28796..28877)  
/rpt\_family="L2"  
complement(29165..29844)  
/rpt\_family="L1PAL10"  
30044..30247  
/rpt\_family="MER3"  
complement(30450..30617)  
/rpt\_family="MIR"  
complement(31814..32035)  
/rpt\_family="L1ME1"  
complement(32043..32073)  
/rpt\_family="(GAAA)n"  
complement(32093..32136)  
/rpt\_family="L1MB7"  
complement(32137..32433)  
/rpt\_family="AluSx"  
complement(32434..32520)





Funding  
Agency : Deutsche Forschungsgemeinschaft (DFG).  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 55470  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/chromosome="2"  
/map="3542416-3597884"  
2458. 6183  
/note="orf\_id:dd\_03152"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43736.1"  
/db\_xref="GI:21166119"  
/translation="MEDEETRLNERRRIRERKKOEEDADKEDQERKAREERY  
RKEARTKEREQAQKKELEYIKRTEERORERERREERKEEKOQKDEERK  
QOERERREDEERKRIKELESRRKVLKELEDSSQVDSILSOKRRLIAR  
KIESILASVDSNOYOTOPTIKIYSLILDELKEVIDLASGGRSIADWSOP  
ASISMAEYLDIVGATIFASKEPFLNGTASLEKNSLAKMSQASELLKGIOTAK  
IITDPSNFIIDILPSMISNPPILPILNMHLFTDVCYSKRKPELLICNNKDTIL  
PSLMQHSNSETVDLLKMAEIMETQSNETSNSGSAASFEFLFTGSSSSSS  
NNNNNSNNKKEPMTNDLTPYLILRINEISLNVIVDQLEIDNORLMEFLVK  
PKRSVAHLGCGSFNAVLNRFKALVHYTSPFSSMLICTLVDSSVTMOSL  
ITNSSSSSSSLSSSSSSSLSSSSSSNODLPVILFLKOIYITDDISNGNSR  
ELSPKASELITIKDSOEFTSYRLKHLISLXKANSKIDREWEGLIKKCI  
DIFREYENNVIRHCIBETLPLIOLRCGSDERPECHLLDCSPKSVNVLGSGN  
SELQINNSNN  
OQSEPVIGAPAPAMGSLSRGGKIRKRVSEFTDEDTSNADENNTSNTGADV  
LAIDLSEFOPPTPAVATTEQSSDKNSLENTVYSEKRFATLSTPSSSSSS  
SSTSTSTSTTTKINCTSSSSSSSGSSISKAELKEOEKOKIKELMDTKSKO  
LTPDOKAISOIKOLKEREKKEKKLARKLEKEKESKSSSTSSSSSSSSSS  
GSVLPSTNKLKSSSSSSSPSVLSTSTTSLAATTTTTPTPTTPTTTTTT  
TWTTTTSLSDSSDTSNSDCSNDSVSELSKLSLNDNSOPTIASVNIESTIE  
QSVVYVHNESSSNNGSNNTCENNSKINDSSTPTSTTATKKGVAHNGHII  
SLCIEITHVNSKGOALRIDYDECCLFGQGDMEVYVKLQEIISFRNKIGTVSMR  
KGRGAPPTPEANESFONMANGVIPPTNPSKFF"

CDS

CDS

CDS

CDS

/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43739.1"  
/db\_xref="GI:21166122"  
/translation="MTLASISSIGNKVSIRKKNVSSSSSSSSSSSSSSSSSS  
GPLTGNVGNLVGLIGTVIGVSGVLTGVTGTPILHPSGCGH"  
join(12494. 12506,12587. 12843)  
/note="orf\_id:dd\_01374"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43740.1"  
/db\_xref="GI:21166123"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
SPLIGVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
complement(join(13453. 13706,13780. 13792))  
/note="orf\_id:dd\_01373"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43741.1"  
/db\_xref="GI:21166124"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
PLIGTVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
join(14479. 14491,14572. 14825)  
/note="orf\_id:dd\_01372"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43742.1"  
/db\_xref="GI:21166125"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
PLITGVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
complement(join(14976. 15348,15627. 15685,15765. 16167))  
/note="orf\_id:dd\_01370"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43743.1"  
/db\_xref="GI:21166126"  
/translation="MAVLTDSLTLNIGSMILLITPILGPIINYYKKKTNPPTPA  
LIVTGKELVGLGIALEAKRTNNSVLGLTGNIKLEIEICUDLGQVETESI  
DVTKEKLYDMLIFEDNKYKIDILIANAGVETMLPELDETERILVTNTNVTGIN  
TLPVNPVFEBSRGSGOLVMSIIPYEDYTAGYSSGSKSGGLIRNLASRGV  
VSVILPGVYPLVDSLDYKIDIPMLPTPTGKAKYIVDGISRDALISCPPLVYCS  
HFAFTPTPLDANFINLSIAKYPDKRPYTHSHHSRSTNSKRENCSSSVKML  
NSQIKITHNK"  
complement(join(16864. 17173,17523. 17530))  
/note="orf\_id:dd\_01367"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43744.1"  
/db\_xref="GI:21166127"  
/translation="MLFSLKKNFINYLFFNIKLPISITIGNKVSIRKSNFSSLSN  
SSLQSSNSIQCGCGGPLIGTVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNC  
GCH"  
join(17865. 17877,17959. 18215)  
/note="orf\_id:dd\_01388"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43745.1"  
/db\_xref="GI:21166128"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
NSLIGTVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
complement(join(18637. 18888,18961. 18973))  
/note="orf\_id:dd\_01385"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43746.1"  
/db\_xref="GI:21166129"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
SPLIGVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
join(19666. 19678,19759. 20015)  
/note="orf\_id:dd\_01382"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43747.1"  
/translation="MADEDDGLQINICLYLOHPVKKKEINSKAEAEAYFRENISNNNFT  
NNRAYREKTNKNN  
YNNNSNN  
IESNNNEKDKYFAENKTKDEENRPLETIESNIYKESFESMMWGSILSETL  
VRNLVGMKHEKPHIOEASTPTPLKGNDAVYKQTSKSLSLIPLYOVKLTQRYT  
RSDCYCIVITPTRELSQITEELQKLPYTVVPGIIMGENRSARAKIRGINI  
LVATPGRLLDHLQNTQSPDNINIKCLIDEDKLLDLGFERDVTIINLDSKRJK  
FKRONIIVSATLSGISRLASLSLSPVYIGLDSKVLKEGPNFOAAEKEMQLPKOL  
DOFVEVESKERLTLIAFIKWTNSITIDGVAAGSSNNSKAIYFESSCSDVDH  
HYMFSNMKMDKERGVKRTKKKQIKODKLIOHKQNSKIOTGNSDESDSDNDSD  
SDSNSSETDEKEIKOIKKATSNRYRKTISVSPYIFKLGLDQOTRTTFD  
FKNSPGLILTTDVSAAGLDPSVNMVVOVDPDCDTPDYIHRIGRTARLQOGCSLIF  
LLEPEKRYIDLAFNYSVEMKYTTLLIYSDQLKTSKSQLESQVHDLQL  
FERLILDSAKEMARCAIOSFLRSYATHKADVYSIFHISYILGHVSKSFLARETPT  
ELNKSAGIKGAKAKGDEDTLOKOTADPFKMYKSVNERSDGLINDOPORANFHE  
SASGVHINIYORHNRRNREDAKNSPKETKEKRRRLYDKKISNNKFKSK"  
complement(join(10317. 10573,10659. 10671))  
/note="orf\_id:dd\_01384"

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43749.1"  
/db\_xref="GI:21166129"  
/translation="MTLASISSIGNKVSIRKSNFSSLSNSSLQSSNSIQCGCGG  
SPLIGVGNLVGLVGTGIIIGTVGVNGVVGGLSGPNCGCH"  
complement(join(19666. 19678,19759. 20015)  
/note="orf\_id:dd\_01382"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAM43747.1"  
/translation="MADEDDGLQINICLYLOHPVKKKEINSKAEAEAYFRENISNNNFT  
NNRAYREKTNKNN  
YNNNSNN  
IESNNNEKDKYFAENKTKDEENRPLETIESNIYKESFESMMWGSILSETL  
VRNLVGMKHEKPHIOEASTPTPLKGNDAVYKQTSKSLSLIPLYOVKLTQRYT  
RSDCYCIVITPTRELSQITEELQKLPYTVVPGIIMGENRSARAKIRGINI  
LVATPGRLLDHLQNTQSPDNINIKCLIDEDKLLDLGFERDVTIINLDSKRJK  
FKRONIIVSATLSGISRLASLSLSPVYIGLDSKVLKEGPNFOAAEKEMQLPKOL  
DOFVEVESKERLTLIAFIKWTNSITIDGVAAGSSNNSKAIYFESSCSDVDH  
HYMFSNMKMDKERGVKRTKKKQIKODKLIOHKQNSKIOTGNSDESDSDNDSD  
SDSNSSETDEKEIKOIKKATSNRYRKTISVSPYIFKLGLDQOTRTTFD  
FKNSPGLILTTDVSAAGLDPSVNMVVOVDPDCDTPDYIHRIGRTARLQOGCSLIF  
LLEPEKRYIDLAFNYSVEMKYTTLLIYSDQLKTSKSQLESQVHDLQL  
FERLILDSAKEMARCAIOSFLRSYATHKADVYSIFHISYILGHVSKSFLARETPT  
ELNKSAGIKGAKAKGDEDTLOKOTADPFKMYKSVNERSDGLINDOPORANFHE  
SASGVHINIYORHNRRNREDAKNSPKETKEKRRRLYDKKISNNKFKSK"  
complement(join(10317. 10573,10659. 10671))  
/note="orf\_id:dd\_01384"

/db\_xref="GI:2116130"  
 translation="MTLLATSSIGNKSISKNFSLSNSSLQSSNSIQCGCGGG  
 SPLTGVNLYGVGIVGTGIIVGVGVGVGVGGLSGNCCG"  
 complement{join(20170..20736,20806..20864,20943..21345)}

```
/product="Putative Oxidoreductase"  
/protein_id="AAM43748.1"  
/db_xref="GI:21166131"
```

translocation="M6VLDKIDLSLLTNIIGSITLLIFPIELGFLNYKKKKINFEPTPA  
IVTGASSGSGRGTALEYAKRRKNSSVYGLGNRNEKLEQECGLGQVQVEESI  
DYWDKEKLDNMLKLPKNQKIKDILLANAGTIEPTLIPKLDPTERTLRYANTNYGTLIN  
TVLDPVPEFSRSGSGQVYVASSITPEPFDVWAGVSSSSGYSIKSGDILNRLASRGV  
VSVLTPPEIIVPVSGLSDYDKDITPMLPIITPDASKSYIVDGIISGALLIISPLLALYIS  
HFAVATIPPTLRDQATNFIINSLLKTPDYKSPVTHYSHSQYSTNESKREDCCPSKAMS  
NSQKSNKAMA"

21759. .22802 |  
/note="ORF\_id: d4\_01376"  
/codon\_start=1  
/product="CYSTEINE-RICH, ACIDIC INTEGRAL MEMBRANE PROTEIN  
PRECURSOR"  
protein\_id="AA043749.1"

Query Match	8.0%;	Score 40;	DB 2;	Length 55470;
Best Local Similarity	51.7%;	Pred. No..3.8;		
Matches 91;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0;

38 AAGCATAAATTTACTTGATTGAATTCACATTGCTGTACATAAATTAGATTGCA 97  
||| ||||| ||| ||| | ||||| ||| |||||  
25935 AATTACCAAATTAATTATAATATAAATTAATTTAAATTCACAATATATAATTTAATTTAA 25964

98 TTTTCGTGTCACCAAGTCCATCCCAATTCATTCGACGCCAATCCGTATATG 157  
||| | | ||| | | | | | | | |  
25995 TTTTATTTCACCAATTTATTATAATTAAATCAATATATTCAGTGAATAATATAAATG 26054

158 GTGCAACATACATGAGTATCTTTCACCAAGAACAGTTAGACCAAGTCCCTTT 213  
 26055 GTAAAAAGTTTATTCATTTGATCCATCAACATCATTTTAAAGAAATAGTCATTTT 26110

Accession	Source	Length (bp)	Notes
U01946	Human	15095	Linear
U01947	Human	15095	Linear
U01948	Human	15095	Linear
U01949	Human	15095	Linear
U01950	Human	15095	Linear
U01951	Human	15095	Linear
U01952	Human	15095	Linear
U01953	Human	15095	Linear
U01954	Human	15095	Linear
U01955	Human	15095	Linear
U01956	Human	15095	Linear
U01957	Human	15095	Linear
U01958	Human	15095	Linear
U01959	Human	15095	Linear
U01960	Human	15095	Linear
U01961	Human	15095	Linear
U01962	Human	15095	Linear
U01963	Human	15095	Linear
U01964	Human	15095	Linear
U01965	Human	15095	Linear
U01966	Human	15095	Linear
U01967	Human	15095	Linear
U01968	Human	15095	Linear
U01969	Human	15095	Linear
U01970	Human	15095	Linear
U01971	Human	15095	Linear
U01972	Human	15095	Linear
U01973	Human	15095	Linear
U01974	Human	15095	Linear
U01975	Human	15095	Linear
U01976	Human	15095	Linear
U01977	Human	15095	Linear
U01978	Human	15095	Linear
U01979	Human	15095	Linear
U01980	Human	15095	Linear
U01981	Human	15095	Linear
U01982	Human	15095	Linear
U01983	Human	15095	Linear
U01984	Human	15095	Linear
U01985	Human	15095	Linear
U01986	Human	15095	Linear
U01987	Human	15095	Linear
U01988	Human	15095	Linear
U01989	Human	15095	Linear
U01990	Human	15095	Linear
U01991	Human	15095	Linear
U01992	Human	15095	Linear
U01993	Human	15095	Linear
U01994	Human	15095	Linear
U01995	Human	15095	Linear
U01996	Human	15095	Linear
U01997	Human	15095	Linear
U01998	Human	15095	Linear
U01999	Human	15095	Linear
U02000	Human	15095	Linear

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC119464				
AC119464.3	GI:21746719			
HTG; HTGS_PHASE1.				
Norway rat.				
Rattus norvegicus				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 150950)

## THOKS

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbra, J., Benton, J., Blinaghe, K., Blankenburg, K., Bonin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Burch, P., Burch, P., Burkett, C., Burnett, K.L., Byrd, N.C., Carron, T., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedecich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Dublin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flisgy, N., Ford, J., Foster, P., Fretzt, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Harrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hultky, S., Hume, Y., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S.,

JOURNAL TITLE	Direct Submission
JOURNAL REFERENCE	Unpublished 2 (bases 1 to 150950)
AUTHORS	Morley, R.C.
TITLE	Direct Submission
JOURNAL	Submitted (27-APR-2002) Human Genome Sequencing Center, Department of

## REFERENCE

JOURNAL

**COMMENT**

Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C., Kratochvíč, J., Kunesil, A., Landry, N., Leal, B., Lewis, L., Li, Y., Li, Z., Licharge, O., Lieu, C., Liu, X., Liu, W., Louissegh, H., Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massery, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogun, M., Okunolu, G., Orangun, N., Ovioldo, R., Pace, A., Payton, B., Peetey, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rotubokan, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shoostrati, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Tellrod, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission

Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 1509550)

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20387289.

Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

```
Center project name: GUXH
Center clone name: CH230-329N4
----- Summary Statistics
```

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 102806 bases at least Q40  
Consensus quality: 108058 bases at least Q30  
Consensus quality: 111642 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.bsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.bsc.bcm.tmc.edu/docs/genbank/draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 49 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of 'N', but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

*	1	1015:	contlig of 1015 bp in length
*	1016	1115:	gap of unknown length
*	1116	2166:	contlig of 1051 bp in length
*	2167	2266:	gap of unknown length
*	2267	4049:	contlig of 1783 bp in length
*	4050	4149:	gap of unknown length
*	4150	5284:	contlig of 1135 bp in length
*	5285	5384:	gap of unknown length
*	5385	6410:	contlig of 1026 bp in length
*	6411	7250:	gap of unknown length
*	6511	7724:	contlig of 1214 bp in length

```

* 7725 7824: gap of unknown length
* 7825 8909: contig of 1085 bp in length
* 7826 9009: gap of unknown length
* 8910 10316: contig of 1307 bp in length
* 9010 10416: gap of unknown length
* 10317 11815: contig of 1399 bp in length
* 10417 11915: gap of unknown length
* 11816 12930: contig of 1015 bp in length
* 11916 13030: gap of unknown length
* 12931 14163: contig of 1133 bp in length
* 13031 14263: gap of unknown length
* 14164 15477: contig of 1214 bp in length
* 14264 15577: gap of unknown length
* 15478 16774: contig of 1197 bp in length
* 15577 16875: gap of unknown length
* 16775 18689: contig of 1815 bp in length
* 16875 18790: gap of unknown length
* 18690 19943: contig of 1154 bp in length
* 18790 20043: gap of unknown length
* 19944 21825: contig of 1782 bp in length
* 20044 21925: gap of unknown length
* 21826 24664: contig of 2739 bp in length
* 21926 24764: gap of unknown length
* 24665 26464: contig of 1700 bp in length
* 24765 26564: gap of unknown length
* 26465 28906: contig of 2342 bp in length
* 26565 29006: gap of unknown length
* 28907 30706: contig of 1700 bp in length
* 29007 30806: gap of unknown length
* 30707 32901: contig of 2095 bp in length
* 30807 33001: gap of unknown length
* 32902 34709: contig of 1708 bp in length
* 33002 34710: gap of unknown length
* 34710 36834: contig of 2025 bp in length
* 34810 36934: gap of unknown length
* 36835 38997: contig of 2063 bp in length
* 36935 39097: gap of unknown length
* 38998 40547: contig of 1450 bp in length
* 39098 40647: gap of unknown length
* 40548 42211: contig of 1564 bp in length
* 40648 44211: gap of unknown length
* 42212 44556: contig of 2145 bp in length
* 44213 44556: gap of unknown length
* 44557 46143: contig of 1587 bp in length
* 44557 46243: gap of unknown length
* 46144 48213: contig of 1970 bp in length
* 46244 48313: gap of unknown length
* 48214 50730: contig of 2417 bp in length
* 48314 50830: gap of unknown length
* 50731 54283: contig of 3453 bp in length
* 50831 54383: gap of unknown length
* 54284 56628: contig of 2245 bp in length
* 54384 60026: gap of unknown length
* 56629 60126: contig of 3298 bp in length
* 56729 60127: gap of unknown length
* 60127 63431: contig of 3305 bp in length
* 60127 63531: gap of unknown length
* 63432 66526: contig of 2995 bp in length
* 63532 66527: gap of unknown length
* 66527 70433: contig of 3807 bp in length
* 66527 70533: gap of unknown length
* 70434 73802: contig of 3265 bp in length
* 70534 73902: gap of unknown length
* 73803 77596: contig of 3694 bp in length
* 73903 77597: gap of unknown length
* 77597 81063: contig of 3367 bp in length
* 77697 81064: gap of unknown length
* 81064 84042: contig of 2879 bp in length
* 81064 84043: gap of unknown length
* 84043 87397: contig of 3255 bp in length
* 84143 87497: gap of unknown length
* 87398 93038: contig of 5541 bp in length
* 87498 93138: gap of unknown length
* 93039

```

```

* 93139 97625: contig of 4487 bp in length
* 97626 97725: gap of unknown length
* 97726 103046: contig of 5321 bp in length
* 103047 103146: gap of unknown length
* 103147 109008: contig of 5862 bp in length
* 109009 109108: gap of unknown length
* 109109 115424: contig of 6316 bp in length
* 115425 115524: gap of unknown length
* 115525 123717: contig of 8193 bp in length
* 123718 123818: gap of unknown length
* 123818 137814: contig of 13897 bp in length
* 137815 137815: gap of unknown length
* 137815 150950: contig of 13136 bp in length.
* 150950: Location/Qualifiers
  1. 150950
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-329N4"

BASE COUNT 41088 a 30687 c 30302 g 41906 t 6367 others
ORIGIN
Query Match 8.0%; Score 40; DB 2; Length 150950;
Best Local Similarity 54.9%; Pred. No. 3.8;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 24 AATTTTCAAGCTAAGTCATAAATATTACTTGAATTCAGTTCATTATGTTTACA 83
Db 124565 AACTTAAATTTTATTAAATATTATTTTATTATGACCCACACATGTTTGGACTCCA 124506

QY 84 TAAATTAGATTGATTTCGTTCTCCAGTCATCCATTCATTCATTCATTCGACGC 143
Db 124505 TAAACGATTCATTCCTTCCTTCTAATTCCTCCATTCCTCCCTACTTCTCTCC 124446

QY 144 AATTCCTTATTTGGTCATACAT 167
Db 124445 ATTGTACGCTTTTCGTTTCTT 124422

RESULT 14
AC112117/c 103138 bp DNA 1linear HTG 17-JUL-2002
LOCUS Rattus norvegicus clone CH230-163M8, *** SEQUENCING IN PROGRESS
DEFINITION Rattus norvegicus clone CH230-163M8, *** SEQUENCING IN PROGRESS
ACCESSION AC112117
VERSION AC112117.3 GI:21744681
KEYWORDS HTG: HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 103138)
AUTHORS Muzny,D.M., Adams,C., Aglio-Oduola,B., All-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaritunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbarta,J., Benton,J., Blinige,K., Blankenburg,K., Bryant,N.P.,
Bourck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Cattoli,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornig,F., Howard,S., Huber,J., Huijks,S., Hunne,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

```

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louie, S.,  
Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Meheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Notchman, N., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogih, M., Okwuon, G.,  
Ogunyeye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pichens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Saverly, G.,  
Shen, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

\*\*\*\*\*  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
\*\*\*\*\*

2 (phases 1 to 103138)  
Worley, K.C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (phases 1 to 103138)  
Worley, K.C.  
Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 13, 2002 this sequence version replaced gt:20303335.

\*\*\*\*\*  
COMMENT  
\*\*\*\*\*

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----  
Center project name: GSIM  
Center clone name: CH230-163M8

----- Summary Statistics -----  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.990329  
Consensus quality: 66968 bases at least Q40  
Consensus quality: 71073 bases at least Q30  
Consensus quality: 75068 bases at least Q20

\*\*\*\*\*  
NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 51 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1048: contig of 1048 bp in length  
\* 1049 1148: gap of unknown length  
\* 2167 2266: gap of unknown length  
\* 2267 3582: contig of 1316 bp in length  
\* 3583 3682: gap of unknown length  
\* 3683 5380: contig of 1698 bp in length  
\* 5381 5480: gap of unknown length  
\* 5481 6512: contig of 1032 bp in length  
\* 6513 6612: gap of unknown length  
\* 6613 7718: contig of 1106 bp in length  
\* 7719 7818: gap of unknown length  
\* 7819 9222: contig of 1404 bp in length

9223 9322: gap of unknown length  
\* 9323 10587: contig of 1264 bp in length  
\* 10587 10687: gap of unknown length  
\* 10687 11789: gap of 1003 bp in length  
\* 11789 11790: gap of unknown length  
\* 11790 13270: contig of 1381 bp in length  
\* 13270 13271: gap of unknown length  
\* 13271 14280: contig of 1019 bp in length  
\* 14280 14380: gap of unknown length  
\* 14380 16666: contig of 2277 bp in length  
\* 16666 16767: gap of unknown length  
\* 16767 18524: contig of 1757 bp in length  
\* 18524 20624: gap of unknown length  
\* 20624 20724: gap of 2000 bp in length  
\* 20724 21833: contig of 1110 bp in length  
\* 21833 21933: gap of unknown length  
\* 21933 23469: contig of 1536 bp in length  
\* 23469 23570: gap of unknown length  
\* 23570 25119: contig of 1450 bp in length  
\* 25119 25120: gap of unknown length  
\* 25120 26242: contig of 1122 bp in length  
\* 26242 26342: gap of unknown length  
\* 26342 27703: contig of 1361 bp in length  
\* 27703 27802: gap of unknown length  
\* 27802 28901: contig of 1099 bp in length  
\* 28901 29001: gap of unknown length  
\* 29001 30885: contig of 1884 bp in length  
\* 30885 30986: gap of unknown length  
\* 30986 32367: contig of 1382 bp in length  
\* 32367 32467: gap of unknown length  
\* 32467 33716: contig of 1249 bp in length  
\* 33716 33816: gap of unknown length  
\* 33816 35447: contig of 1631 bp in length  
\* 35447 35547: gap of unknown length  
\* 35547 36882: contig of 1335 bp in length  
\* 36882 36982: gap of unknown length  
\* 36982 38392: contig of 1410 bp in length  
\* 38392 38493: gap of unknown length  
\* 38493 40161: contig of 1669 bp in length  
\* 40161 40261: gap of unknown length  
\* 40261 41932: contig of 1671 bp in length  
\* 41932 42032: gap of unknown length  
\* 42032 43343: contig of 1311 bp in length  
\* 43343 43444: gap of unknown length  
\* 43444 44841: contig of 1398 bp in length  
\* 44841 44941: gap of unknown length  
\* 44941 46760: contig of 1819 bp in length  
\* 46760 46860: gap of unknown length  
\* 46860 47948: contig of 1088 bp in length  
\* 47948 48048: gap of unknown length  
\* 48048 49790: contig of 1742 bp in length  
\* 49790 49890: gap of unknown length  
\* 49890 51345: contig of 1455 bp in length  
\* 51345 51445: gap of unknown length  
\* 51445 53769: contig of 2324 bp in length  
\* 53769 53869: gap of unknown length  
\* 53869 55228: contig of 1360 bp in length  
\* 55228 55329: gap of unknown length  
\* 55329 58219: contig of 2790 bp in length  
\* 58219 58220: gap of unknown length  
\* 58220 60515: contig of 2296 bp in length  
\* 60515 60616: gap of unknown length  
\* 60616 62620: contig of 2005 bp in length  
\* 62620 62720: gap of unknown length  
\* 62720 63883: contig of 1163 bp in length  
\* 63883 63984: gap of unknown length  
\* 63984 65312: contig of 1329 bp in length  
\* 65312 65412: gap of unknown length  
\* 65412 68650: contig of 3238 bp in length  
\* 68650 68750: gap of unknown length  
\* 68750 71059: contig of 2309 bp in length  
\* 71059 71060: gap of unknown length









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - protein search, using frame\_plus.n2p model

Run on: July 4, 2003, 19:02:57 ; Search time 913 Seconds

(without alignments)  
13505.226 Million cell updates/sec

File: US-10-083-053b-2  
Percentage score: 52328  
Sequence: 1 gatatgtaagaagcctca.....caattctgtgaagaagtaa 29921

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xih  
-O=/cgn2\_1/USPTO.spool/US1008353/runatc03072003\_093611\_8089/app\_query.fasta.1.30087  
-DB=SPREMBL\_21 -OPT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cgi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USBR=US1008353 @cgn\_1.1549 @runatc03072003\_093611\_8089 -NCPU=6 -ICU=3  
-NO\_MAP -LARGEBOUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WANN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database 8  
1: SPREMBL\_21  
2: SP\_dacteria:  
3: SP\_fungi:  
4: SP\_human:  
5: SP\_invertebrate:  
6: SP\_mammal:  
7: SP\_mhc:  
8: SP\_organelle:  
9: SP\_phase:  
10: SP\_plant:  
11: SP\_podent:  
12: SP\_virus:  
13: SP\_vertebrate:  
14: SP\_unclassified:  
15: SP\_virus:  
16: SP\_bacteriap:  
17: SP\_archaeap:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match Length	ID	Description
1	5202	9.9	1275 4	000370

## SUMMARIES

2	5195	9.9	1275	4	015604	015604 homo sapien
3	5190	9.9	1275	4	000360	000360 homo sapien
4	5188	9.9	1275	4	09Y560	09Y560 homo sapien
5	5188	9.9	1275	4	09UN80	09UN80 homo sapien
6	5186	9.9	1275	4	08TE30	08TE30 homo sapien
7	5182	9.9	1275	4	000366	000366 homo sapien
8	5182	9.9	1275	4	000375	000375 homo sapien
9	5176	9.9	1275	4	015606	015606 homo sapien
10	5175	9.9	1275	4	000363	000363 homo sapien
11	5169	9.9	1275	4	012881	012881 homo sapien
12	5168	9.9	1275	4	000378	000378 homo sapien
13	5163	9.9	1275	4	000362	000362 homo sapien
14	5150	9.8	1275	4	000368	000368 homo sapien
15	5133	9.8	1275	4	000372	000372 homo sapien
16	3605.5	6.9	1281	11	09QW12	09QW12 mus muscullu
17	3605.5	6.9	1281	11	054850	054850 mus muscullu
18	3600.5	6.9	1281	11	088913	088913 mus muscullu
19	3597	6.9	1300	11	P97692	P97692 rattus norv
20	3595.5	6.9	1281	11	088915	088915 mus muscullu
21	3594.5	6.9	1281	11	09QWY3	09QWY3 mus muscullu
22	3590.5	6.9	1300	11	008906	008906 mus muscullu
23	3589.5	6.9	1281	11	008910	008910 mus muscullu
24	3588.5	6.9	1281	11	088914	088914 mus muscullu
25	3559	6.8	1275	6	062658	062658 canis fami
26	3558.5	6.8	1281	11	091289	091289 mus muscullu
27	3557.5	6.8	1281	11	060713	060713 mus muscullu
28	3557.5	6.8	1281	11	091288	091288 mus muscullu
29	3533.5	6.8	1300	11	061785	061785 mus muscullu
30	3511.5	6.7	1252	11	09QWY2	09QWY2 mus muscullu
31	3366.5	6.4	1219	11	09QWY3	09QWY3 mus muscullu
32	2883.5	5.5	712	4	014754	014754 homo sapien
33	2653.5	5.1	641	4	014288	014288 homo sapien
34	2284.5	4.4	573	4	000549	000549 homo sapien
35	2237	4.3	611	4	014288	014288 homo sapien
36	2234	4.3	1275	4	000362	000362 homo sapien
37	2234	4.3	1275	4	000363	000363 homo sapien
38	2230	4.3	1275	4	000366	000366 homo sapien
39	2229	4.3	1275	4	015604	015604 homo sapien
40	2229	4.3	1275	4	09UN80	09UN80 homo sapien
41	2229	4.3	1275	4	015606	015606 homo sapien
42	2226	4.2	1275	4	09Y560	09Y560 homo sapien
43	2225	4.2	1275	4	000378	000378 homo sapien
44	2224	4.2	1275	4	000370	000370 homo sapien
45	2222	4.2	1275	4	000375	000375 homo sapien

## ALIGNMENTS

RESULT 1  
ID 000370 PRELIMINARY; PRT; 1275 AA.  
AC 000370;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative p150.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
ON NCBI\_TaxID=9606;  
RX MEDLINE=97285120; PubMed=9140393;  
RA Sasanian D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,  
RA Deberardinis R.J., Gabriel A., Svergold G.D., Kazazian H.H. Jr.;  
RT "Many human l1 elements are capable of retrotransposition";  
RL Nat. Genet. 16:37-43(1997).  
DR EMBL: U93569; AAC51271.1; -;  
DR InterPro: IPR005135; Exo\_endo\_phos.  
DR InterPro: IPR000477; RVTse.  
DR Pfam: PF03372; Exo\_endo\_phos; 1.  
DR Pfam: PF00078; lvt; 1.  
KW RNA-directed DNA polymerase.

SEQ SEQUENCE 1275 AA; 149011 MW; 588703688E7129FF CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 1275  
Score: 5202.00 Matches: 1059  
Percent Similarity: 79.18% Conservative: 78  
Best Local Similarity: 73.75% Mismatches: 133  
Query Match: 9.94% Indels: 173  
DB: 4 Gaps: 7  
US-10-083-853b-2 (1-29921) x 000370 (1-1275)  
QY 3428 ATGGCAGGATCAAAATTCACACATTAATATTAACCTTAATGTAATGGGCTAAATTC 3487  
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnSer 20  
QY 3488 CCAATTAAGACACAGACTGCAAAATGGATTAAGCTCAAGACCCATAGCTGTCT 3547  
DB 21 ProlIleYsArgHisArgLeuAlaSerThrPileYsSerGlnsProlSerValCysCys 40  
QY 3548 ATTCCAGAGGCCCTCTACATGAAAAAGACACATAGGCTCAAAATTAAGGATGGAG 3607  
DB 41 IleGlnGluThrHisIleuThrCysArgAspThrHisArgLeuYsIleYsGlyTyrArg 60  
QY 3608 AAGATTACCACTAAATGGAACAAAAAAGACAGGGCTGCATCTAGCTCT 3667  
DB 61 LysIleTyrGlnIleAsnGlyLysGln---LysLysAlaGlyValAlaIleValSer 79  
QY 3668 GATTAACAGACTTAAACCAAGATCAAAAGACAAAGGACCTTAATCAATAG 3727  
DB 80 AspLysThrAspPheLysProlThrLysIleLysArgSpLysGlnGlyHisTyrIleMet 99  
QY 3728 GTAAGGCAATGGAACAAAGAGCTAACTATCTTAATATAGATGACCACTAATCA 3787  
DB 100 ValIleGlySerIleGlnGlnGlnIleuThrIleLeuAsnIleTyrAlaProlAsnThr 119  
QY 3788 GGGAGCCAGATTCATAAGCAAGTCTTAGAGACCTTACAAAGACTTGGACTCCAC 3847  
DB 120 GlyAlaProlArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139  
QY 3848 ACAATATAGTGGAGCTTAATTAATATAGACACTTAAACCCCACTGCCAATATTA 3907  
DB 140 ThrLeuIleMetLys-----AspPheAsnThrProlSerIleLeu 153  
QY 3908 GGCAGATCAATGAGACAGAAATTAACAGGATATCCAGAGAGTGAATGAGCTGGAC 3967  
DB 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173  
QY 3968 CAAGCGACCTAATATATATCTACAGACTCCCACTCCCAATCAACAGAAATTAACATC 4027  
DB 174 GlnThrAspLeuIleAspIleTyrArgThrLeuHisProlYsSerThrGluTyrThrPhe 193  
QY 4028 TTCTGACATCACTATACACCTATTTAAATGACCATGATATTTAGTAACAACTC 4087  
DB 194 PheSerAlaProlHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213  
QY 4088 CTCAGCAATGCAAAAAGAACAGAAATCTAAACAAAGTCTCTCAGACTACAGTGCATC 4147  
DB 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233  
QY 4148 TATTTGAACCTAGAAATTAAGAACTCACTCAAAATCACAACAATCATGGAAGTGAAC 4207  
DB 234 LysLeuGlnLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn 253  
QY 4208 AACCTGCTCTGTAATAGTACTAGTGGTAAATTAACAAATGAAGGCAAAATTAAGATGTC 4267  
DB 254 AsnLeuLeuLeuAsnSpLysTyrTrpValHisAsnGluMetLysAlaGluIleYsKerPhe 273  
QY 4268 TTTGAACCAATGAGAAACAAGACACATATACCAAAATCTCTGGGCAATATTATTAAGA 4327  
DB 274 PheGluThrAsnGlnLysAspThrThrTyrGlnAsnLeuThrPsrAlaPheYsAla 293  
QY 4328 GTGTGTAGAGGCAAAATTTATAGCACTAGATGCTTACAGAGAAAGCAAGAAATATCTAAA 4387

DB 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313  
QY 4388 ATGACACCTTAATACATACATTAATAAGAACTGAGAAAGAAAGCAAACTTAATCAAAA 4447  
DB 314 IleAspThrLeuThrSerGlnIleLysGlnLeuGlnLysGlnIleGlnThrHisSerLys 333  
QY 4448 GCTAGCAGAAAGACAAATAACTAATAGATCAGACAGAACTGAAGAGATAGACAA 4507  
DB 334 AlaSerArgArgGlnGlnIleThrLysIleArgAlaGluLeuYsGluIleGluThrGln 353  
QY 4508 AAAGCCCTTAATTAATCAATGATCAGAGACCTGGTTTGGAAAGTCAGCAAAAT 4567  
DB 354 LysThrLeuGlnLys-IleAsnGluSerArgSerThrPheGlnAlaArgIleAsnLysIle 373  
QY 4568 ----AGACCACTAGACACTAATTAAGAAAGAAAGAGAGAAATCAAAAGACATCAAT 4623  
DB 373 eAspArgProlLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle 393  
QY 4624 AAAAATGATTAAGGGGATTCACACCCAGATCCACAGAAATCAAACTATTATCAGAGA 4683  
DB 393 eLysAsnAspLysGlyAspIleThrThrAspProlGlnIleGlnThrThrIleArgGlu 413  
QY 4684 ATATTTAACACCTCTATGCAATTAATACTAGAAATCTAGAAAGAAATGGATTAATCTCT 4743  
DB 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGlnLysGlnIleMetAspThrPheLe 433  
QY 4744 GGACACATATGTAGCTGTATGACCTTGGGGGACAGAAACAAAGGGGGTGAATGCAGAA 4803  
DB 433 uAsp----- 434  
QY 4804 ATAAAGACAAAGACAAAGAGTATGTTGGAAAGTGAAGGGTCAGGGGCACTTGCTCT 4863  
DB 434 ----- 434  
QY 4864 AATGCAAGAGGCCCTGAGCTTTACACACACCTCTGATTTATTAGCAAAAGATAGAC 4923  
DB 434 ----- 434  
QY 4924 GAGAGGTGATGGAAGAAAGAGTGAAGTGAAGTCCAGAGTGAAGCTGCAGAGACTGC 4983  
DB 434 ----- 434  
QY 4984 ATTCTCAAAACAATAGGCTTAGATGTCCAGATAGATTAACCTCAAGAGCCAGTGCAGG 5043  
DB 434 ----- 434  
QY 5044 GAGTGAATGGCCCTCAGCAAAACCTTGAAGGACAGGACAGAAAGTAAGTTGCCACATTCCT 5103  
DB 434 ----- 434  
QY 5104 GTATTCAGATTAACAGATTGCTGTTTGATCAAGTAAGTCCAGTGAATGCTGAGTGG 5163  
DB 434 ----- 434  
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCAAAAACATCAACACCTCTCAGACTAAACCA 5223  
DB 435 ----- 435  
QY 5224 GGAAGAGTCAATCCCTGAATATACCACTAACCAATGTTTAAATTAAGACAGTATTTGA 5283  
DB 443 nGluGlnValGlnSerLeuAsnArgProlIleThrGlySerGluIleValAlaIleLeuAs 463  
QY 5284 TAGCTACCAACCAAAAAAGTCCAGAGCAGAGGATTCACAGCAAAATTTTACACAGAG 5343  
DB 463 nSerLeuProlThrLysLysSerProlGlyProlAspGlyPheThrAlaGluPheTyrIleArg 483  
QY 5344 GTACAAAGAGAAGCTGGTACTATTCCTTGTGAATATTCACAAAATATGA---AATGG 5400  
DB 483 gTyrLysGlnGluLeuValProlPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503  
QY 5401 GAATCTCCCTAATCTATTTACAGAGCCAGCATCTCTGATACCAAACTAGACAGTC 5460  
DB 5401 ----- 5401

```
Dh 503 lylleleuProasnSerPheTyrGlnAlaSerIleIleuIleProIysProGlyArgA 523
Oy 5461 ACACAACAAAAGAGAAATTCAGGCCCATATCCCGATGAACTTGGATGAAATCC 5520
Dh 523 sphrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleI 543
Oy 5521 TCATATAAAATCTGGCAAAACCAATCCAGCAGCATCAAAAGACTTATCTACATGATC 5580
Dh 543 eunsInLysIleleuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563
Oy 5581 AATTTGGCGTCATCCCTGGGATGCGAGCGTGTCAAAATATGCAATCAATCAATAGTAG 5640
Dh 563 lvalGlyPheIleProGlyMetGlnIlyTrpPheAsnIleArgLysSerIleAsnValI 583
Oy 5641 GCCATCATATAAGAACCAATGATGACAAAACCAATGATTTATCTCAATGATGACAGAAA 5700
Dh 583 lcglnHisIleAsnArgAlaLysAspLysAsnHisValIleIleSerIleAspAlaLul 603
Oy 5701 AGGCGCTTTGTCAAAATTCACACGCCCTTCATGCTAAAAATTCATCAATCAATGATTCG 5760
Dh 603 lylAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623
Oy 5761 ATGAATGATCTCTCAAAATATAAGAGCTATTATAC-AAACCCACAGCCAAATATCTAC 5819
Dh 623 spEllyMetLysLysLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleI 643
Oy 5820 TGAATGGGCAAAAACGTGAGACATTCCTTGAGAACTGGCACAGACAGAGATGCCCTC 5879
Dh 643 eunsInGlyLysLysGlnAlaPheProLeuLysThrGlyIlyTrpArgGlnGlyCysProL 663
Oy 5880 TCTCACACACTCTATTCACAACTATTTGGAAGTTCTGGCCAGGCAATCAGGCAATAGA 5939
Dh 663 euserProLeuLeuPheAsnIleValLeuGlnValleuAlaArgAlaIleArgGlnIul 683
Oy 5940 AAGAATATAAGGCTATTCAATAGAAAGAGAGAGAGATCATTTGCTCTGTTGCAGATG 5999
Dh 683 lylGlnIleLysGlyIleGlnLysGlnLysGlnLysLysLeuSerLeuPheAlaAspA 703
Oy 6000 ACATGTTGTATATTGAAACCCCATGCTCAGGCCCAAAAACCTCAAGCTGATTA 6059
Dh 703 spMetIleValIlyrLeuGlnAsnProIleValIleSerAlaGlnAsnLeuLysLeuIleS 723
Oy 6060 GCAACTTCACAAAGCTCTCAGACACCAAAATCAATGCAAAAATCAACAAGATTTCTAT 6119
Dh 723 ersnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT 743
Oy 6120 ACCGCCATATATAGCAAAACAGAGAGCCAAATCATGAGTCACTCTTCACAATGCTTA 6179
Dh 743 lylAsnAsnAsnArgGlnThrGlnSerGlnIleMetGlyLysLeuProPheThrIleAlaS 763
Oy 6180 CAAGAAGAAATAAATACCTAGATACACTTCAAGAGACAGTACAGACTCTTCAAG 6239
Dh 763 eLysArgIleLysIlyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
Oy 6240 AGAAGTCAAAACACATGATCAAGAAATAAGAGAGACACAAACAAATGAAAAACATTC 6299
Dh 783 lunsInLysLysProLeuLeuLysGlnIleLysGlnLysAspThrAsnLysTrpLysAsnIleP 803
Oy 6300 CAGTCTCACAGATAGTAGAATCAT-----GAAAATGCCATATCTGCCAAAGTAAT 6352
Dh 803 rocLysSerTrpValGlyArgIleAsnIleValIlyLysMetAlaIleLeuProLysValIleT 823
Oy 6353 ATGATATACGTGTACCCCATCAGTACATTCATCTTCTCACAGATTTGAAAAA 6412
Dh 823 lylArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGlnLeuGlnLysT 843
Oy 6413 CAACATTAAATTTCAATAGAACCAAAAAAGAGCCACAGACAGCAAGCAATCTTAAGC 6472
Dh 843 hrThrLeuLysPheIleTrpAsnGlnLysArgAlaArgIleAlaLysSerIleLeuSer 862
Oy 6473 AAAAAAGCAAAAGCTGGAGATATCATGCTTCACTTAAACTATATCTTAAGGCTACA 6532
Dh 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuLysLysAlaThr 882
Oy 6533 GTAACCAAAACCTGATGCTGCTGATACCAAAACAGATATATACCAATGAGACAGAAC 6592
Dh 883 ValThrLysThrAlaTrpIlyrTrpGlnAsnArgAspIleAspGlnTrpAsnArgThr 902
Oy 6593 GAGACCTCAGAAATTC-ACATGCAATCTACATCATCTGATCTTTGCAAAACCTGACAAA 6651
Dh 903 GluProSerGlnIleMetProHisIleTyrAsnTrpIleuIlePheAspLysProGlnLys 922
Oy 6652 AACACAGCAATGAAAAAGATTCCTCTTATTAATATGCTTTGAAAAACCTGCTAGCC 6711
Dh 923 AsnLysGlnTrpGlyLysAspSerLeuLeuAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
Oy 6712 ATATGCAAGAGCTGAAATGATGATCCCTTCATACACCTTATCAAAAGTAATCAAGA 6771
Dh 943 lIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
Oy 6772 TGAATTAAGACTTAAATATATAGACATTAACCAATAAAACCA-GAAGAAAACCTAGGC 6830
Dh 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnGlnAsnLeuGly 982
Oy 6831 AATACCATTCAGATATGACATGCGCAAGTCTCATGACTTAAACACCAAAACCAATG 6890
Dh 983 lIleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
Oy 6891 GCAACAAACCCAAATATGACAGTGGATCTGATTTAACTATAGACTCTGCACAGCA 6950
Dh 1003 AlathrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
Oy 6951 AAAAAAACCTGTCATCAGACTGGAACAACAACACTACAGAAATGGAGAAAAATTTTGCAT 7010
Dh 1023 -LysGlnThrThrIleArgValAsnArgGlnProThrThrTrpGlnLysIlePheAlaTh 1042
Oy 7011 CTATGATCTGACAAAGCTAATATATCCAGAGATCTACGAAAGACTTAAACAAATTTTACA 7070
Dh 1042 llyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleLysLys 1062
Oy 7071 GAAAAA-----ACACCCCGCTCAAAATATGCGGAAGATATGACAGACACTTCTCAAA 7126
Dh 1062 slLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgIlePheSerLys 1082
Oy 7127 AGAAGACATTTATGACACCAACAACAATATGAAAAAACCTCATCATGATGGTCTTAG 7186
Dh 1082 sgLysAspIleTyrAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleArg 1102
Oy 7187 AGAAATGCAAAACAAACCAAGTGCATATCCATCTCATCTGATTTGAATGGTGTAC 7246
Dh 1102 gGlnMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI 1122
Oy 7247 TAAAAAGTCAAGAAACAAACAAATGCTGAGAGATGGAGAAATAGAACACTTTTCCA 7306
Dh 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGlnIleGlyThrLeuValHI 1142
Oy 7307 CTGTGGTGGGAATGTAATTAATTAAGTCAACCATTTGTGAGACAGTGTGAGATCTTAA 7366
Dh 1142 scLysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 1162
Oy 7367 GGATCTTGAAACCGAAATATCATTTTGACCCAGAAATCCCATTTACTGAGTATATACCAA 7426
Dh 1162 gAspLeuGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI 1182
Oy 7427 GGAATTAATTAATCATTTATTAATTAAGACATGACATGATGATTTATTCACAGACAT 7486
Dh 1182 sAspTrpLysSerCysCysLysLysAspTrpTrpArgMetPheIleAlaAlaLeuPn 1202
Oy 7487 CACAATAGCAAAAGCTTTGAAACCAACCAATGTCATCATGATGATGATGATGATGATAAGAA 7546
Dh 1202 eThrIleAlaLysThrTrpAsnGlnProAsnGlyCysProThrMetIleAspTrpIleLysLys 1222
Oy 7547 AACATGGCAATATACCAATGAAATATCTATGAGGCAATTAAG- GATGAGTTCATGTC 7605
Dh 1222 sMetTrpHisIleLysThrThrMetGlnLysLysLysLysLysLysLysLysLysLysLysLys 1242
```

QY 7606 CTTTGACAGAGATATGATGAGAGCTGGAAACCATCATCTTCAGCAAACTAACAAGACA 7665  
DB 1242 rpevalglutThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGlu 1262  
QY 7666 GAAACCAACACACACATGTTCTCAGCTGTAAGTGGAGT 7705  
DB 1262 nlvsthrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275  
RESULT 2  
Q15604  
ID Q15604 PRELIMINARY; PRT; 1275 AA.  
AC Q15604;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE CREF2 contains a reverse transcriptase domain.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dombroski B.A.;  
RT "Isolation of an active human, transposable element.";  
RL Science 0:0-0(1991).  
DR EMBL; M80340; AAAS1622.1; -  
DR InterPro; IPR005135; Exo\_endo\_phos.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF03372; Exo\_endo\_phos; 1.  
DR Pfam; PF00078; rvt; 1.  
KM RNA-directed DNA polymerase.  
SQ SEQUENCE 1275 AA; 149044 MW; C090473BB0A44C1 CRC64;

## Alignment Scores:

Pred. No.: 0 Length: 1275  
Score: 5195.00 Matches: 1057  
Percent Similarity: 78.97% Conservative: 77  
Best Local Similarity: 73.61% Mismatches: 136  
Query Match: 9.93% Indels: 173  
DB: 4 Gaps: 7

US-10-083-853b-2 (1-29921) x Q15604 (1-1275)

QY 3428 ATGGCAGATCAATATTCACATATTAATTAATTAATTAATTAATTAATTAATTAATTC 3487  
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnGlnGlyLeuAsnSer 20  
QY 3488 CCAATTAAGACACAGACGTCGCAATGATTAAGATCAAGACCATCAGTGTCTGT 3547  
DB 21 AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValLysCys 40  
QY 3548 ATTACAGAGGCCATCTCATATGAAAAAGACACATAGGCTCAAAATTAAGGATGAGG 3607  
DB 41 IleGlnGluThrIleSerThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60  
QY 3608 AAGATTACCACTAATTAAGAAAAAAGAAAAAAGAGGGTGTCAATCTAGTCTCT 3667  
DB 61 LysIleTyrGlnAlaAsnGlnLysGln--LysLysAlaGlyValAlaIleLeuValSer 79  
QY 3668 GATTAACAGACTTTAAACCAACAAAGATCAAAAGAGACAAGAGCCATTACATTAAG 3727  
DB 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGlnLysIleLysIleMet 99  
QY 3728 GTAAGGACATCAATGACAAAGAGAGCTAACTATCTTAATTAATTAATTAATTAATTA 3787  
DB 100 ValLysGlySerIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119  
QY 3788 GGAGCACCACGATTCATAAAGCAAGTCTTAGAGACCTACAAAGAGACTTTGACTCCAC 3847  
DB 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLysGlnArgLysLeuSerHis 139  
QY 3848 ACAATTAATAGTGGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3907  
DB 111

DB 140 ThrIleLeuMetGly-----AspPheAsnThrProLeuSerThrLeu 153  
QY 3908 GGCAGATCAATGACAGAGAAATTAACAAGAGATATGACAGAGTGAATGACCTGTGAC 3967  
DB 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnIleLeuAsnSerAlaLeuHis 173  
QY 3968 CAAGGAGCCTAATTAAGATATCTACAGAACTCCCAACCCCAATCAACAGATTAACATC 4027  
DB 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGlyThrPhe 193  
QY 4028 TTTCAGCATCATATTAACCTTTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4087  
DB 194 PheSerLysArgHisIleThrLysIleAspHisIleValIleLysSerLysAlaLeu 213  
QY 4088 CTCAGCAAAATGCAAAAGACAGAAATCTTAACAAGAGTCTCAGACTACATGCAATC 4147  
DB 214 LeuSerLysCysLysArgThrGlnIleIleThrAsnThrLysLeuSerHisSerAlaIle 233  
QY 4148 TATTTAAGCTCAGATTAAGAAATCTACTCAAAATCAACAATCAATGAAATGAAAC 4207  
DB 234 LysLeuGlnIleuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn 253  
QY 4208 AACCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4267  
DB 254 AsnLeuLeuLeuAsnAspLysTrpValHisAsnGlnMetLysAlaGlnIleLysMetPhe 273  
QY 4268 TTTGAACCAATGACAGAAACAGACACAAATCTCAGGAGCATATTAAAGCA 4327  
DB 274 PheGlnThrAsnLysAsnLysAspThrThrThrGlnAsnLeuThrAspAlaPheLysAla 293  
QY 4328 GTGTGTAGAGGAAATTTATAGCATAGATGCTTACAAAGAGAAAGCAAGAAATTTGAAA 4387  
DB 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGlnLysSerLys 313  
QY 4388 ATGACACCTTAATCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4447  
DB 314 IleAspThrLeuThrSerGlnLysGlnLeuGlnLysGlnLysGlnLysGlnLysGln 333  
QY 4448 GCTAGCAGAAAGACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4507  
DB 334 AlaSerArgArgGlnGlnIleThrLysIleArgAlaGlnIleuLysGlnIleGlnThrGln 353  
QY 4508 AAGCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4567  
DB 354 LysThrLeuGlnLys-IleAsnGlnSerArgSerTrpPhePheGlnArgIleAsnLys 373  
QY 4568 ----AGACACCTGACAGACTAATTAAGAAAGAAAGAGAGAGAAATCAAGATGCAAT 4623  
DB 373 AspArgProLeuSerArgLeuIleLysLysLysArgGlnLysAsnGlnIleAspThr 393  
QY 4624 AAAAAATGATTAAGGGATATCACACCGATCCACAGAAATTAATTAATTAATTAATTAAT 4683  
DB 393 LysAsnAspLysGlyAspIleThrThrAspProThrGlnIleGlnThrIleArgGln 413  
QY 4684 ATTTTAAACACCTTATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 4743  
DB 413 LysIleTyrLysHisLeuTyrAlaAsnLysLeuGlnLysGlnLysGlnLysGlnLys 433  
QY 4744 GGACACATATGTAGCTGTATGACCTGTGGGGACAGAACAAAAGGGGTGAATGACAGAA 4803  
DB 433 unsp----- 434  
QY 4804 ATTAAGACAAAGACAAAGAGATGTTTGAAGTAGGGGTCAGGGGCAACTTGCTCT 4863  
DB 434 ----- 434  
QY 4864 AATGCAAGGGCCCTGAGCTTTACACCAACCTCTGTATTATTATAGCAAAAGAGATAGC 4923  
DB 434 ----- 434  
QY 4924 GAGAGGCTAGTTGGAAGAAGAGCTGAGCTGTAGCTCAGAGTAGGCTGCAAGACTGC 4983  
DB 434 ----- 434

OY 4984 ATTCTCAAAACAATAGGCTGTAGATGTCCACAGTAACTCAAGAGACCGCAGTCCAGG 5043  
Db 434 ----- 434  
OY 5044 GAGTATGAGCCCTCAGCAAACTCTAGGCGACAGCAGAGTAAGTTGCCACATTCT 5103  
Db 434 ----- 434  
OY 5104 GTATTACAGATAAAGATTGCTGTTTGTATCAAGTAGCCTCCAGTGGATGCTGAGTTGG 5163  
Db 434 ----- 434  
OY 5164 TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACATATACCCCTCTCAGACTAAACCA 5223  
Db 435 -----thrythrileuProargleuasn1 443  
OY 5224 GGAAGAGTCAAAATCCCTGCAATATACCAATAACAAGTTCTAAATTTGAAGCAGTAATTGA 5283  
Db 443 nglugluValgluserleuasnArgProilethrghysergluileValalailleleas 463  
OY 5284 TGGCTTACCAACCAAAAAAGTCCAGACAGCAGATTCACAGCCAAATTTCTACAGAG 5343  
Db 463 nserleuProthrLysLysSerProglProaspglPheThrAlaGluphetyGlnArg 483  
OY 5344 GTACAAAGAGAGCTGTACTATCTCTTGAAGACTATCCAAAAATAGAA---AATGG 5400  
Db 483 gtrymetglugluValleuValProphleuLeuLysleuPheGlnSerilleglulysglu-G 503  
OY 5401 GAATCCCTCCCTACTCATTTTACGAGGCGCAGCATCCTGTATACCAAAACCTAGCAGTGG 5460  
Db 503 lyilleuProasnserPheTyrgluAlaserilleleuileProLysProglYArgA 523  
OY 5461 ACACAACAAAAAGAAATTTACAGCCCATATCCCTGTATGAAACATGTATGGAANAATCC 5520  
Db 523 sPTHrLysLysgluasnPheArgProileserleuMetasnilleasPalalysileL 543  
OY 5521 TCAATAAAATACTGGCAAAACCAAAATCCAGCAGCATCAAAAAGTTATCTACCATATC 5580  
Db 543 euasnLysilleuValaasnArgilleglnGlnHisleLysLysleuileHisHisasg6 563  
OY 5581 AAGTTGGCGTCATCCCTGGGATGCAAGCGTGTCAAAATATGCAAAATCAATTAATGTAG 5640  
Db 563 llnValglYpHeilleProglmetGlnGlyTrpPheasnilleArgLysSerilleasValil 583  
OY 5641 GCCATCACATAAAGAACCATGACAAACCAACCATGATTTATTCATAAGATGCGAAG 5700  
Db 583 leGlnHisLleasnArgAlaasnAspLysasnHisMetilleSerilleasPalagluL 603  
OY 5701 AGCGCTTGTCAAAATTCACAGCCCTTCATGCTAAATAATTTCTAGTAAGTATGATCG 5760  
Db 603 ysAlaPheaspLysilleglnGlnProPheMetleuLysThrleuasnLysleuGlyila 623  
OY 5761 ATGGAATGTATCTCAAAATATATAGAGCTATTTATAC-AAACCCACAGCCCAATATCATAC 5819  
Db 623 spglYThrLysrPheLysilleleArgAlaileTyraspLysProthrAlaasnilleleL 643  
OY 5820 TGAATGGGCAAAACGGAAGCATGCCCTTGAGAACTGGCACAGAGAGAGATGCCGTC 5879  
Db 643 euasnGlyLyslnLysleuGlnAlaPheProleuLysThrleYThrArgGlnGlyCysProL 663  
OY 5880 TCTCACCACCTCTATTCAGATACTATTTGGAAGTTCTGGCCAGGCAATCAGCAGATAGA 5939  
Db 663 euSerProleuLeuPheasnilleValleuGlnValleuAlaArgAlaileArgGlnGluL 683  
OY 5940 AAGAAATAAAGGATTTCAATAAGAAAGAGAGAGAGTCTATTTGCTCTGTTTGCAGATG 5999  
Db 683 ysglnilleLysglYilleGlnleuGlyLysgluGlnlValalysLeuSerleuPheAlaAspA 703  
OY 6000 ACATGTTTATATTATAGAAAACCCCATGCTCAGGCGCAAAAACCTCTTAAGCTGATGA 6059  
Db 703 spMetilleVallyrleuGlnuasnProilleValSerAlaGlnasnleuLysleuLies 723

OY 6060 GCAACTTCAGCAAGTCTCAGACACAAAATCAATGTGCAAAAATTCACAGCATTCCTAT 6119  
Db 723 eRAsnPheSerLysValserGlyTyrlYsilleasnValGlnLysSerGlnAlaPheleu 743  
OY 6120 ACCCAATATATAGCAAAACAGAGAGCCAAATCATAGTGAATCTTCATTCACAAATTCGA 6179  
Db 743 yThrAsnAsnArgGlnThrGlnSerGlnilleMetGlygluLeuProPheAlilleas 763  
OY 6180 CAAGAGATTAATAATACCTAGGAATPACAATTCACAGGACACGTAGCAACCTTCACAG 6239  
Db 763 erLysArgilleLyslyrleuGlyilleGlnleuThrArgaspAlalysaspLeuPheLysG 783  
OY 6240 AGAAGTACAAACACCTGATCAAGAAATTAAGAGAGACACAAACAAATGGAANAATTC 6299  
Db 783 llnasnTyrlsProleuLeuLysgluileLysgluLnsPheThrLysstrpLysasnilleP 803  
OY 6300 CATGCTCAGCATAGTAAGATCAT-----GAAAATCCATCTGCCCAAGTAATTT 6352  
Db 803 rocysSerTrpValglArgilleasnilleValLysMetAlaileleuProLysValilleT 823  
OY 6353 ATAGATTCAGTGTACCCCATCAAGCTACCATGTACTCTTCACAGAAATGGAAAAA 6412  
Db 823 yThrPheasnAlaileProilleLysleuProMetThrPhePheThrGluLeuGluLysT 843  
OY 6413 CAACTTTAATTTATATATGAAACCAAAAAAGCCACAGACGCAAGCAATCTTAACG 6472  
Db 843 hThrleuLysPheilleThrPasnGln-LysArgAlaArgilleAlalysSerilleLeuSer 862  
OY 6473 AAAAAGACAAAGCTGGAGGATTCATGCTGACTTAACATATATATAGGCTTACA 6532  
Db 863 GlnLysasnLysAlaGlylylleThrleuProaspPheLysleuTyrlYsAlaThr 882  
OY 6533 GTAACCAAAACGTGAGTACGTGTACCAAAAACAGTATATAGCAAAAGGAACGAACA 6592  
Db 883 ValThrLysThrAlaThrPtyrTrpTyrglnasnAlaArgaspLleaspLlnThrPasnArgThr 902  
OY 6593 GAGACTCAGAAATTTAC-ACGTCAATCTCATCATCATGATCTTTTGACAAACCTGCACAA 6651  
Db 903 GluProserGlnilleMetProHislellyrAsnTyrlleuilePheaspLysProglulys 922  
OY 6652 AACCAAGCAATGCAAAAAGGATTCCTATTTAATTAATGGTGTGGAAAACTGGCTAGCC 6711  
Db 923 AsnLysGlnTrpGlyLysaspSerleuPheasnLysTrpCysTrpLlnasnTrpLeuAla 942  
OY 6712 ATATGAGAAAGCTGAAACTGATCCCTCTTACACCTTATACAAAATTAACCTCAAGA 6771  
Db 943 llecysArgLysleuLysleuasnProPheleuThrProTyrlThrLysilleasnSerArg 962  
OY 6772 TGAATTAAGACTTAATATATAGACATAAACCATAAACCACCA-GAAGAAAACCTAGGC 6830  
Db 963 TrpilleLysaspLeuasnValLysProLysThrilleLysThrleuGlnGlnasnleuGly 982  
OY 6831 AATACCATTCAGATATATGACATGGGCAAAAGCTTCATGACTTAAACACCAAAAGCATG 6890  
Db 983 llerThrilleGlnAspilleGlyAlaGlyLysaspPheMetSerLysThrProLysAlaMet 1002  
OY 6891 GCAACAAAAGCCAAATATAGCAAGTGGGATCGATTAACCTATAGAGCTTCGACAGAGA 6950  
Db 1003 AlaThrLysaspLysilleasnLysstrpAspLeuileLysleuLysSerPheYThrAla 1022  
OY 6951 AAAAAAACTGTATCATCAGATGAGACAGCAAGCAACCTACAGAAATGGAGAAAAATTTGCAAT 7010  
Db 1023 -LysGlnThrThrilleArgValAlasnArglnProThrThrTrpGluLysillePheAlaThr 1042  
OY 7011 CTATGATCTGACAAAAGGCTAATATCCAGAGATCTPACGAAGAATTAACCAATTTACAA 7070  
Db 1042 rTyrlSerSerAspLysGlyLysleuileSerArgilleTyrasnGlnleuLysGlnilleTyrlY 1062  
OY 7071 GAAAAA-----AACACCCCGTCAAAAATATAGGCAAAAGATATAGCAGACACTTTCAAA 7126  
Db 1062 slYsLysThrAsnAsnProilleLysLysstrpAlaLysaspMetAlaArgHisPheSerly 1082  
OY 7127 AGAAGACATTTATGACAGCCACAAACATATGAAAAAACCTTCATCATCTTGGTGTAG 7186

```
Db 1082 sgluasplletryraalaalalsylshismetlylsysscyserserleuallailear 1102
Oy 7187 AGAATGCAAAAACACACAGATGATACCATCTGCTAGTAAATGGTGATAC 7246
Db 1102 gclumetclnlelystrthrmelargtyrhishleuthrprovalaigmelalailell 1122
Oy 7247 TAAAGTCAGAAACAACAATGCTGAGAGATGTGAGAAATAGAACACTTTTCCA 7306
Db 1122 elysylsserserlysnasnarcgystprargglycysglglnileglythrleuLeuH 1142
Oy 7307 CTGTGCGTGGAAATGTAATAGTTCACACATGTGGAGACAGTGTGAGATTCCTTAA 7366
Db 1142 scystprtrpaspyslyslsleuValGlnproleuthrlysserValtrpargphleuar 1162
Oy 7367 GGATGTAGAACCCAGAAATATGATTTGACCCAGCAATCCATTAAGTATGAGTATACCCAA 7426
Db 1162 gasplreulnleuengulileprophesaprhoalalerleuLeuGlylletryprohs 1182
Oy 7427 GGAATATAATCATTCATTTATTAAGACACATGCACACATATGTTTATGACAGACTGAT 7486
Db 1182 nglutrylsserserCysTyltrylsasprthrcysrthargmethrheleallaaleuPh 1202
Oy 7487 CACATATGCAAGCTTGGAAACCAACCAATGTCACATGATGATGATGATGATGATGAT 7546
Db 1202 ethrleallalysrthtrpansglnprolyscysprothrmethleasprtlelysl 1222
Oy 7547 AACATGGCAGATATACACATGAAATACATGACAGCCATAAAG-GATGAGTTCATTC 7605
Db 1222 smetrphtsiletryrthrmelglutryrlyalalailelysnasaprgluhneillese 1242
Oy 7606 CTTTGACAGATATGATGATGAAAGCTGGAACCATCATCTCAGCAAACTACACAGAACA 7665
Db 1242 rPheValglythrtrpmetlyslsleuGlnthrlelleuSerlyslsleuSerGlnlgl 1262
Oy 7666 GAAACCAACACCACTGCTCTCACTGTAAAGTGGSGGT 7705
Db 1262 nlystrlyshisargliepheserleuileglycylasn 1275

RESULT 3
ID 000360 PRELIMINARY: PRT: 1275 AA.
AC 000360.
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUN-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deberardinis R.V., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human l1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL: U93563; AAC51261.1;
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RYase.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; Ivt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149062 MW; 350B4F0358E525F0 CRC64;
```

```
US-10-083-853b-2 (1-29921) x 000360 (1-1275)
Oy 3428 ATGCGAGATCAAAATTCACACATATAATATTAACCTTAATGTAATGGCTAAATTC 3487
Db 1 methrlyserasenseniillethrilleuthrleuasnilleasnGlyleuasnser 20
Oy 3488 CCAATTAAGACACAGACTGCAAAATGATTAAGATCAAGACCCATGATGTGCTGT 3547
Db 21 AlailelysarGHisargargAlaserprlleysserGlnasprSerValCysCys 40
Oy 3548 ATTCAGAGGCCCATCTCACTGAAAGACACATAGGCTCAAAATTAAGGATGGAGG 3607
Db 41 lleGlnluthrHisleuthrCysrArgasprthrsalargleuLysllelyGlytrarg 60
Oy 3608 AAGATTTTCAAGTAAATGCAAAACCAAAAAAAGAGGAGGTTGCAATCTAGTCTCT 3667
Db 61 lyltletryGlnAlaasnGlylsGln---lyslsalaglyValAlaileuValser 79
Oy 3668 GATAAACAAGACTTTAAACAACAAGATCAAAAGACAAAGAGCCATTACATAATG 3727
Db 80 AsplystrhrasrphelyprothrllysileysargasrlysglnGlythlewet 99
Oy 3728 GTAAAGCATCAATGGAACAAGAGAGCTACTATCTCTAAATATATGACCCCAATGCA 3787
Db 100 VallysglyserlleGlnlglnglnlgluthrthlleuasnilletryrAlaproasnthr 119
Oy 3788 GGAGACCCAGATTCATTAAGCAAGTCTTGAAGACCTTAAGACAAAGAGACTTGGACTCCAC 3847
Db 120 GlylarproalrphellelyslnglnValleuSerasrpleuGlnalrghsleuAsprserHis 139
Oy 3848 ACAATATATGTTGGAGTCTAAATATTAATAGACACTTAAACCCCATCAATATTA 3907
Db 140 ThrleullemetGly-----AsprheasnthrProleuSerthrleu 153
Oy 3908 GCGCATCAATAGACACAGAAATTAACAAGATATCCAGAGTGAACGTGACTGTGAC 3967
Db 154 AsphargserthrargGlnlyslValasnlysnasprthrglnlgleuasnserAlaileuHis 173
Oy 3968 CAAGGAGCATTAATAGATATATACAGAACTCCCAACCCCAATCAACAGATATATACATC 4027
Db 174 GlnAlasrpleuileasprlletryrArgthrleuHisprolysserthGlntrythrPhe 193
Oy 4028 TTTCAGCATCACAATACCACTTAATTAATTAATGACATGTAATTTAATGAACACTC 4087
Db 194 PheleuAlaproHisHisrthlyserlylleasprHisleValglyserlyAlaleu 213
Oy 4088 CTCAGCAATGCAAAAGACAGAAATCTCTAACAAACAGTCTCTCAGACTACAGTGCATC 4147
Db 214 leuSerlyscyslysarGlythrGlnllethrtrasnlyrleuSerAsprHisserAlaile 233
Oy 4148 TATTAGAACTCAGAAATTAAGAAACTCACTCAAAATCAACAATACATGGAACCTGAC 4207
Db 234 lysleuGlnleuarglielelysnleuthrlnseralnargserthrthtrlylyleuasn 253
Oy 4208 AACCTGCTCTGAAATGACTACTGGTAAATATACAAATGAAAGCAAAATTAAGATGTC 4267
Db 254 AsnleuLeuLeuasnAsprlytrValHisnGlnleuLyssalaglnllelymeprhe 273
Oy 4268 TTTGAACCAATGGAACAACAAGACAAATGACCAAGATCTCTGGGGATTTTAAGCA 4327
Db 274 PheGlnluthrnsnglnuasnlysnasprthrthlyrGlnsnleuthrparAlaPheLysla 293
Oy 4328 GTGTGTAGAGGAATTTATATGACTGATGCTTACAAAGAGAAAGCAAGAAATATCTAAA 4387
Db 294 ValCysargglylysrphelleleleuasnAlaItrylsarglysglnGlnIargSerlyls 313
Oy 4388 ATAGACACCTTAACATCACAATTAAGAAGACTAGAGAAGAAAGCAAAATTCAAA 4447
Db 314 llesprhrleuthrserGlnleuLygluenglnlysglnglnGlnthrHisserlyls 333
Oy 4448 GCTAGCAGAAAGACAGAAATTAATACATCAGACACAGAACTGAAGGAGATGAGACACA 4507
```



```
Db 334 AlaSerArgArgIngluilethrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
OY 4508 AAGCCCTCAAAATTAATCAATGAATCCAGAGCTGGTTTGGAAAAGATCGACAAAT 4567
Db 354 LysThrLeuGlnLys-IleasnGluSerArgSerTrpPheGluArgIleasnLysIl 373
OY 4568 ----AGACCACTAGACACTAATTAAGAAAGAGAGAGAGATCAAGAGATGCAAT 4623
Db 373 eaSPArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleasprHIl 393
OY 4624 AAAAATGATAAAGGGGATATCAACACCGATCCACAGAAATCAAACTATATACAGA 4683
Db 393 elYsAsnAspLysGlyAspIleThrThrAspProthGluIleGlnThrThrIleArgG1 413
OY 4684 ATATTATTAACACCTCTATCAATATAACTAGAAAATCTGAAGAAATGATAAATCCCT 4743
Db 413 uTrpTrpLysHisLeuThrAlaAsnLysLeuGluAsnLeuGluLeuArgIleAsnThrPheLe 433
OY 4744 GGACACATATGTAGCCCTGTATGGACCTTGGGGAGACAGAACAAAAGGGGTGAATGCAGAA 4803
Db 433 uAsp----- 434
OY 4804 ATAAAAGACAAACAAAGAGATATGTTTGGAAAGTACGGGCTCAGGGGCACTTGCCCTCT 4863
Db 434 ----- 434
OY 4864 AATGACAAGGGCCCTGAGCTTACACACCCTCTGTATTTATTAGGCAAAAGAGATAGC 4923
Db 434 ----- 434
OY 4924 GAGAGGGTGAGTTGGAAGAGAGGTCAAGCTGTAGTCCAGAGTACGCTCGACAAGACTGC 4983
Db 434 ----- 434
OY 4984 ATTCCCAACAATAGGCTTAGATGTCCAGTAGATAACTCAAGAGGCCAGTGCAGG 5043
Db 434 ----- 434
OY 5044 GAGTGATGGCCCTCAGCAAACTTCTTAGGGCAGGACAGAAAGTAAGTTGCCACATTCCT 5103
Db 434 ----- 434
OY 5104 GATATCGATAAACAGTTTGTCTGTATGATCAAGTACGCTCCAGTGAATGCTGAGTTGG 5163
Db 434 ----- 434
OY 5164 TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACATACACCTCTCAAGACTAAACA 5223
Db 435 -----ThrTyrThrLeuProArgLeuAsnG1 443
OY 5224 GGAAGAAGTCAATCCCTGAATATACAGTAAAGTCAAAATTTGAACAGCAGTAATGA 5283
Db 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs 463
OY 5284 TAGCCTACCAACAAAAAGATCCAGACAGACAGGATTCACAGCCAAATTCACAGAG 5343
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheYrGlnAr 483
OY 5344 GTCAAAAGAAACCTGCTACTATTCCTTTCGAACATTTCCAAAAAATAAA---AATGG 5400
Db 483 gTyTrpLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503
OY 5401 GAATCCCTCCCTAATCTTTAGAGAGCCAGCATCATCGATACCAAAACCTAGCAGTG 5460
Db 503 LylleuLeuProAsnSerPheTrpGluAlaSerIleIleLeuIleProLysProGlyArgA 523
OY 5461 ACAACAACAAAAGAGAAATTCAGGCGCATATCCCTGATGAACATGATGTGAATAATCC 5520
Db 523 sPThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543
OY 5521 TCATATAAATACGCGCAACCAATCCAGACAGACATCAACAAAAGCTTATATACATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnIleHisIleLysLysLeuIleHisIleAspG 563
OY 5581 AAGTGGCGTCATCCCTGGGATCGAGGCGTGTCTCAAAATATCAATCAATATAGTAG 5640
Db 563 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleAlaGlySerIleAsnValI 583
OY 5641 GCCATCAATAAACAGAACCAATGACAAAACCAATGATTAATCTCAATAGATGACAGAA 5700
Db 583 LegInHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 603
OY 5701 AGGCCTTGTCAAAATTCACACAGCCCTTCATGCTAAAAATTCACGTAATCTAGGTACG 5760
Db 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuYsrThrLeuAsnLysLeuGlyIleA 623
OY 5761 ATGAATGTATTCACAAATATAAGACTTTTATAC-AAACCCACAGCCCAATATATAC 5819
Db 623 spGlyThrTrpPheLysIleIleArgAlaIleTrpAspLysProthAlaAsnIleArgL 643
OY 5820 TGAATGGGCAAAAACCTGGAAGCATTCCTTTGAGAACTGGCACACAGAGATGCCCTC 5879
Db 643 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 663
OY 5880 TCTCACACATCTTATTCACATACTATGGAAGTTCTGGCCAGGCAATAGCAATAGA 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL 683
OY 5940 AAGAAATTAAGGGATTCAAATAGAAAGAGAGAGAGATCATATGTCTGTGGCAGATG 5999
Db 683 yseGluIleLysGlyIleGlnLeuGlnLysLeuGluValLysLeuSerLeuPheAlaAspA 703
OY 6000 ACATGTTGTATATTTAGAAAACCCATCGCTCGAGCCAAAACCTCCTTAAGCTGATA 6059
Db 703 spMetIleValIlyrLeuGlnLysAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIle 723
OY 6060 GCAACTTCAGCAAGTCTCAGACACACAAATCAATGTGCAAAATACAGACTTCTTAT 6119
Db 723 eAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuTr 743
OY 6120 ACCCAATTAATAGACAAACAGAGAGAGCCAAATCATAGTGAATCTCATTACAAATTCCTA 6179
Db 743 yTrhAsnAsnArgInThrGluSerGlnIleMetSerGluLeuProPheThrIleAsAs 763
OY 6180 CAAGAAGATTAATATCTAGGAATACACTTCAAGAGACAGCAGTGAACCTTCAAGG 6239
Db 763 eLysArgIleLysTrpLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
OY 6240 AGAACTTCACACATGATCAAGAAATAGAGAGGACACAAACAAATGTGAATAATTC 6299
Db 783 LuAsnTrpLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP 803
OY 6300 CATGCTCACAGATAGTAAAGATCAT-----GAAAATGGCATATCCCAAGTAATTT 6352
Db 803 rocYsSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleTr 823
OY 6353 ATAGATTGAGTGTACCCCATCAAGCTACATGACTTTCACAGAAATGAAAAA 6412
Db 823 yTrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysTr 843
OY 6413 CAACTTAATTCATATGGAACCAAAAAGAGCCACAGACGCAAGCAATCTTAAGC 6472
Db 843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaLysIleAlaLysSerIleLeuSer 862
OY 6473 AAAAAGAAACACCTGGAGGTATCATGCTACCTGACTTAACATATCTTAAGGCTACA 6532
Db 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTrpLysAlaThr 882
OY 6533 GTAACCAAACTGCATGCTGCTGACCAAAACAGATATATAGCAANTGAAACAGACA 6592
Db 883 ValThrLysThrAlaTrpTrpTrpGluAsnArgspIleAspGlnTrpAsnArgTrh 902
OY 6593 GAGACCTCAAGAAATTAACACT-GCAATCTACATCCATCTGTCTTGAACAAACCTGACAAA 6651
Db 903 GluProSerGluIleThrProHisIleTyrAsnTrpLeuIlePheAspLysProGluLys 922
```

```

QY 6652 AACACAAATGCAAAAGATTCCTTATTATTAATGCTTGGAAAAAGCTGGCTAGCC 6711
    |||||
Db 923 Asnlysglnttrpilylsaspsrleuphasnlystrpocystrpolsntrpaula 942
QY 6712 ATATGAGAAGATGGAATGATCCCTCCTACACCTTATCAAAAGTAACCTGACGA 6771
    |||||
Db 943 IleCysArlylsleuylsleuysppropeleuthProlythrythylleasnsery 962
QY 6772 TGAATTAAAGACTTAAATTAAGACATTAACCATTAACCAAGCAAGCAAGCAAG 6830
    |||||
Db 963 TrpIlelyaspheusnvalargProlystrIlelystrleuglunleugly 982
QY 6831 AATACCATTCAGATATGACATGGCAAGACTTATGACTTAAACCAAGCAAGCAAG 6890
    |||||
Db 983 lIethrIleglinsplleeglythrglylsaspPheMetSerlystrProlysalMet 1002
QY 6891 GCACAAAGCCCAATTAAGACAAAGTGGATCTGATTAACCTATAGAGCTTCTCAGACA 6950
    |||||
Db 1003 AlathrlyslalyslleasplystrpaspheulllelylsleuylsserPheCysThrAla 1022
QY 6951 AAAAAAAGCTGTCAGAGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 7010
    |||||
Db 1023 -lysglnttrpIleargValasnarlglnProthrtrtrpGlutyllePhehAlath 1042
QY 7011 CTAATCATGTCAGAAAGCTTAATATCCAGAGATCTACGAGAACTTAAACAAATTACAA 7070
    |||||
Db 1042 rlyrSerSerAspIlysglyleuIleSerArlyrIleArglnleuylsglntIleTyrly 1062
QY 7071 GAAAAA-----AACCAACCCGTCGCAAAATATGGCAAGATAGCAGACACTTCCAAA 7126
    |||||
Db 1062 slslystrhrasnasproIlelysntrpAlalysaspMetasnarHHisPheSerly 1082
QY 7127 AGAAGACATTTATGACGCAACAAACATATGAAAAAACCTCATCATTTGGTCGTAG 7186
    |||||
Db 1082 sglusprIleTyrAlaAlalyslyshsmetylsyScysSerSerleuAlalIleAr 1102
QY 7187 AGAAATGCAAAACAAACCAAGCTGACATACCATCTCATGCTAGTTAGATGGTGCAC 7246
    |||||
Db 1102 gclumetglntIlelystrhrMetArtyrIlestrhrProvalArgMetAlalIle 1122
QY 7247 TAAAAATGAGGAACAAACAAATGCTGAGAGATGAGAAATGAGAACTTTTCCA 7306
    |||||
Db 1122 elyslysserGlyAsnasnarGlystrpArglyCysgllylIleGlyThrleuunH 1142
QY 7307 CTGTTGGTGGGAATGTAATTAAGTTCACCATTTGGAAGACAGTGGAGATTCCCTTA 7366
    |||||
Db 1142 scystrpTrpAspCyslyslleuValglInProleuTrpIlysserValtrpArgPheleuAr 1162
QY 7367 GGATCTAGAACCAAGAAATTCATTTGACCCAGCAACCATTAAGTACAGATATACCCAAA 7426
    |||||
Db 1162 gaspleugluleuGluIlePropeasproAlalIleProleuIleGlyIleTyrProas 1182
QY 7427 GCAATATTAATCATCTTATTAAGACACATGACACATGATGTTATGACAGACTGAT 7486
    |||||
Db 1182 nasprlyrlysserCysCysTyrlylsasprhrCysThrArgMetPheIleAlalaleuAr 1202
QY 7487 CACAATAGCAAAAGCTTGAACCAACCAATGTCATCAGTATGATAGACTGGATAAGAA 7546
    |||||
Db 1202 eThrIleAlalysThrtrpAsnlglnProlyscysprothrMetIleasprIlelysl 1222
QY 7547 AACATGAGGCATATACCATGAATAGTACATGAATGAATGAATGAATGAATGAATGA 7605
    |||||
Db 1222 shetrpIshIleTyrThrMetGluTyrTrpAlaAlalIlelylsasnasplPheMetse 1242
QY 7606 CTTTCAGAGATATGATGACGTAAGCAATCATCTTCACCAAACTAACACAGACAGA 7665
    |||||
Db 1242 rPheValglYthrTrpMetlylsleuGlutThrIleIleValSerlylsleuSerGlnGlu 1262
QY 7666 GAAAAACAAACACACATGTTCTCACTTGAAGTGGAGAT 7705
    |||||
Db 1262 nlystrhrlyshIstargIlePheSerleuIleGlyGlyAsn 1275

```

```

QY95K0
ID QY95K0 PRELIMINARY; PRT: 1275 AA.
AC QY95K0;
DT 01-NOV-1999 (T-REMBLrel. 12, Created)
DT 01-NOV-1999 (T-REMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-REMBLrel. 21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Divoky V., Indra K., Mrug M., Brabec V., Huisman T.H.J., Prechal J.T.;
RT "A novel mechanism of B-Chalassemia. The insertion of L1
RT retrotransposable element into B globin IVSII."
RL Blood 88:148-148(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Kimberland M.L., Divoky V., Prechal J., Schwahn U., Berger W.,
RA Kazarian H.H.;
RT "Full-length human L1 insertions retain the capacity for high
RT frequency retrotransposition in cultured cells."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF149422; ABD38785.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; rvc; 1.
KW Hypothetical protein; RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149015 MW; 7E4AE526015ED37C CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5188.00 Matches: 1056
Percent Similarity: 79.04% Conservative: 79
Best Local Similarity: 73.54% Mismatches: 135
Query Match: 9.91% Indels: 173
DB: Gaps: 7

US-10-083-853b-2 (1-29921) x QY95K0 (1-1275)
QY 3428 ATGGCAGGATCAAAATTCACACATTAATTAATTAATTAATTAATTAATTAATTC 3487
    |||||
Db 1 MetThrGlySerAsnSerHisIleThrIleuThrLeuAsnIleAsnIleuAsnSer 20
QY 3488 CCAATTAAGACACACAGCTGCAAAATGATTAAGACGACAGACCATCAGTGTCTGT 3547
    |||||
Db 21 AlalIlelyAsrGHisArgIleuAlaserttrpIlelysserGlnAsproSerValCys 40
QY 3548 ATTCAGGAGGCCCATCTCACATGAAAAGACACACATAGGCTCAAAATTAAGGATGAGG 3607
    |||||
Db 41 lIeGlnGluThrHisIleuThrCysAsrGpThrHisArgLeuIlylIeysglYtrpArg 60
QY 3608 AAGATTTCACAGTAATTAATGAAAACAAAAAAGCGGGGTGGCAATCCTACTCTCT 3667
    |||||
Db 61 lylsIleTyrGlnAlasnsGlylsGln--lyslsAlaGlyValAlalIleuValSer 79
QY 3668 GATTAAGACGACTTTAAACCAACCAAGATCAAAAGACAAAGAGAGCCATTAACATTAATG 3727
    |||||
Db 80 AsplystrhrAspPheIlyserprothrIlystrIlelysaGpIlysglunlyshIstrIleMet 99
QY 3728 GTAAAGCATCAATGAGAACACAGAGCTTAATCTTAATATATACATGACACCAATACA 3787
    |||||
Db 100 VallysglySerIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119
QY 3788 GGAGCACCACAGATTATATAAGCAAGTCTTAGACAGCTACAAAGAGACTTTAGCTCCAC 3847
    |||||
Db 120 GlyAlaProArgPheIlelysglnValleuSerGlnleuGlnArgAspleuAspserHis 139
QY 3848 ACAATTAATAGTGGGAGTCAATATATATATATATAGACACTTAACACCCGCTGCAATATTA 3907
    |||||
Db 140 ThrleuIleMetGly-----AspPheAsnThrProleuSerThrleu 153

```



QY 3908 GGCAGATCAATGACAGAAAATTAAACAAGGATTCGAGGATTTGAATCGCTGCAC 3967  
||||| ||||||| ||||||| ||||||| |||||||  
Db 154 AsparGSerThrArgGlnLysValAsnLysAspThrGlnGlnLysAsnSerAlaLeuHis 173  
QY 3968 CAAGCGACCTAATAGATATCTACAGAACTCCCAACCCCAATCAACAAGATATACATC 4027  
||||| ||||||| ||||||| ||||||| |||||||  
Db 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGlnTyrThrPhe 193  
QY 4028 TTCTCGACATCAATACACCTATTTTAAATGGACCATGTATTTTAACTAAACATC 4087  
||||| ||||||| ||||||| ||||||| |||||||  
Db 194 PheSerAlaProHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213  
QY 4088 CTCGACAAATGCAAAAGACAGAAATTCCTAAACAAGCTCTCAGACTCAGTCAATC 4147  
||||| ||||||| ||||||| ||||||| |||||||  
Db 214 LeuSerLysCysLysArgThrGlnIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233  
QY 4148 TATTTGAACCTCAAGATTTAAGAACTCCTCAAAATCAACAATCAATGGAAGCTGAC 4207  
||||| ||||||| ||||||| ||||||| |||||||  
Db 234 LysIleGlnIleuArgIleLysAsnLeuThrGlnSerArgSerThrThrTyrLysLeuAsn 253  
QY 4208 AACCTGCTCTGATGACTGCTGGTAAATTAACAAATGAAAGCAAAATTAAGTCTC 4267  
||||| ||||||| ||||||| ||||||| |||||||  
Db 254 AsnLeuLeuLeuAsnAspTyrTyrValHisAsnGlnLysAlaGlnIleLysMetPhe 273  
QY 4268 TTTGAAACCAATGAGAAACAAGACACATGTACAGATCTCTGGGCAATATTAAAGCA 4327  
||||| ||||||| ||||||| ||||||| |||||||  
Db 274 PheGlnThrAsnGlnLysAspThrThrTyrGlnAsnLeuThrAspAlaPheLysAla 293  
QY 4328 GTGTGTAGAGGAAATTTATAGCACTAGATGCTCAAGAGAAAGCAAGAAATATCTAAA 4387  
||||| ||||||| ||||||| ||||||| |||||||  
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnIleuArgSerLys 313  
QY 4388 ATGACACCTTAATCAATCAATTTAAAGAACTAGAGAAAGAAAGCAAAATTCAAA 4447  
||||| ||||||| ||||||| ||||||| |||||||  
Db 314 IleAspThrLeuThrSerGlnLysGlnLysGlnLysGlnIleGlnThrHisSerLys 333  
QY 4448 GCTAGGAGAAAGCAAGAAATTAAGATGACAGACAGAACTGAAAGATAGACAGACA 4507  
||||| ||||||| ||||||| ||||||| |||||||  
Db 334 AlaSerArgGlnGlnIleThrLysIleArgAlaGlnLysGlnIleGlnThrGln 353  
QY 4508 AAAAGCTTCAATTAATCAATGATTCAGAGACTGGTTTTTGAAGATCAAGCAAAAT 4567  
||||| ||||||| ||||||| ||||||| |||||||  
Db 354 LysThrLeuGlnLys-IleAsnGlnSerArgSerThrPheGlnLysIleAsnLysII 373  
QY 4568 ----AGACCACTAGACAGACTAATTAAGAAAGAGAGAGAAATCAAGAGATGCAAT 4623  
||||| ||||||| ||||||| ||||||| |||||||  
Db 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgGlnLysAsnGlnIleAspThrII 393  
QY 4624 AAAAATGATAAAGGGATATTCACACCCAGTCCACAGAAATCAAACTTTATCAGAGA 4683  
||||| ||||||| ||||||| ||||||| |||||||  
Db 393 eLysAsnAspLysGlnAspIleThrThrAspProThrGlnIleGlnThrThrIleArgGln 413  
QY 4684 ATTTTATTAACACCTCTATGCAATTAACATAGAAATCTAGAAAGAAATGATTAATCTCT 4743  
||||| ||||||| ||||||| ||||||| |||||||  
Db 413 utTyrTyrLysHisLeuTyrGlnLysLysLysLysGlnAsnLeuGlnIleMetAspThrPheLe 433  
QY 4744 GAGACATATGTAGCCTGTATGAGACCTTGGGGGACAGAACAAAAGGGGGTGAATGCGAA 4803  
||||| ||||||| ||||||| ||||||| |||||||  
Db 433 uAsp----- 434  
QY 4804 ATAAAGACAAAGACAAAAGATATGTGTTGGAAGTAGGGGTCAAGGGGCAATCTGCCTCT 4863  
||||| ||||||| ||||||| ||||||| |||||||  
Db 434 ----- 434  
QY 4864 AATGCAAGAGCCCTGAGCTTTACACACCCCTGTATTTATTAGCAAAAGAGATAGC 4923  
||||| ||||||| ||||||| ||||||| |||||||  
Db 434 ----- 434  
QY 4924 GACAGGGTGAAGTTGGAAGAGAGGTACAGCTGTAGGTCCAGAGTAGGCTGCAAGACTGC 4983  
||||| ||||||| ||||||| ||||||| |||||||  
Db 434 ----- 434

QY 4984 ATTCTCAAAACAATAGGCTCTAGATGTCCAGTAGATTAACCTTAAGAGGCCAGTGCAGG 5043  
----- 434  
Db 434 ----- 434  
QY 5044 GAGTAGAGCCCTCAGCAAACTTTAGGCGAGGCACAGAAAGTAAGTTGCCACATCTCT 5103  
----- 434  
Db 434 ----- 434  
QY 5104 GTATTCAAGATTAAGAACTTGTGCTTTGATCAAGTAGACCTCCAGATGCGAGAGTTGG 5163  
----- 434  
Db 434 ----- 434  
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCCAAAACATATACACCTCTCAAGATTAACCA 5223  
----- 435  
Db 435 -----ThnTyrThrLeuProArgLeuAsnGln 443  
QY 5224 GGAAGAGTCAATTCCTGAATATACCAGTATACAACTGTAAATTAAGACGTAATTTGA 5283  
||||| ||||||| ||||||| ||||||| |||||||  
Db 443 ngLysGlnValGlnSerLeuAsnArgProIleThrGlnGlySerGlnIleValAlaIleIleAs 463  
QY 5284 TAGCTTACCAACCAAAAAAAGTCCAGAGACAGACGGATTTACAGCCCAATTTACAGAG 5343  
||||| ||||||| ||||||| ||||||| |||||||  
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGlnPheTyrGlnAla 483  
QY 5344 GTACAAAGAGAGAGCTGTAATCTCTTGTGAACCTTTTCCAAAAATAGAA--AATGC 5400  
||||| ||||||| ||||||| ||||||| |||||||  
Db 483 gTyrLysGlnGlnLeuValProPheLeuLeuLysLeuPheGlnSerIleGlnLysGln-u 503  
QY 5401 GAAATCTCCTAACCTATTTTACAGAGCCACATCATCTGATACCAAAACCTAGACAGTG 5460  
||||| ||||||| ||||||| ||||||| |||||||  
Db 503 LysIleLeuProAsnSerPheTyrGlnLysSerIleIleLeuIleProLysProGlyArg 523  
QY 5461 ACACACAAAAAGAGAAATTTTCAAGGCCATATCCCGAGAACATGATGATGAAATCC 5520  
||||| ||||||| ||||||| ||||||| |||||||  
Db 523 sPthrThrLysLysGlnAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543  
QY 5521 TCAATAAATACCTGGCAACCAATCCAGAGACATCAAAAGCTTATCTACCATGATC 5580  
||||| ||||||| ||||||| ||||||| |||||||  
Db 543 eLysLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563  
QY 5581 AAGTGGCGTCATCCCTGGGATGCAAGCGTGGTTCAAAATATGCAATCAATTAATGTAG 5640  
||||| ||||||| ||||||| ||||||| |||||||  
Db 563 lValGlyPheIleProGlnMetGlnLysTyrPheAsnIleArgLysSerIleAsnValI 583  
QY 5641 GCCATCACTAAACAGAACCAATGACAAAAACCAATGATTAATCTCAATGATGCGAGAA 5700  
||||| ||||||| ||||||| ||||||| |||||||  
Db 583 leGlnHisIleAsnAlaArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaLul 603  
QY 5701 AGGCTTTGTCAAAATTCACACAGCCCTTCATGCTAAATTTCTAGTAAACTAGTATCG 5760  
||||| ||||||| ||||||| ||||||| |||||||  
Db 603 ySalArgPheAspLysIleGlnGlnProPheMetLeuLysThrIleAsnLysLeuGlnIleA 623  
QY 5761 ATGGAATGTATCTCAAAATTAAGAGCTATTTATAC-AAACCCACAGCCAAATATCATAC 5819  
||||| ||||||| ||||||| ||||||| |||||||  
Db 623 spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL 643  
QY 5820 TGAATGGCAAAAGCTGAGAGCAATCCCTTGAAGAACTGGCAGACAGCAAGATGCCCTC 5879  
||||| ||||||| ||||||| ||||||| |||||||  
Db 643 eLysnGlnGlnLysLeuGlnAlaPheProLeuLysThrGlnTyrArgGlnGlnLysProL 663  
QY 5880 TCTCACACATCTTATCAAGATATCTATTTGAAGTCTGGCGGACGCAATGACAGCAAPAG 5939  
||||| ||||||| ||||||| ||||||| |||||||  
Db 663 eLysProLeuLeuPheAsnIleValIleuGlnValIleuAlaAlaArgAlaIleArgGlnIul 683  
QY 5940 AAGAAATTAAGGATATTCAAATAGAAAGAGAGAGAAAGTCAATATGTCTCTGTTGCAGATG 5999  
||||| ||||||| ||||||| ||||||| |||||||  
Db 683 ysgLysLysGlnIleGlnLeuGlnLysGlnLysValLysLeuSerLeuPheAlaAspA 703  
QY 6000 ACATGTTGTATATTTTGAAGAAACCCCATGCTTCAGAGCCAAAAATCTTAACCTGATTA 6059  
||||| ||||||| ||||||| ||||||| |||||||  
Db 703 spMetIleValTyrLeuGlnAsnProIleIleIleSerAlaGlnAsnLeuLysLysLeuIleL 723  
QY 6060 GCAACTTCAGCAAAAGTCTCAGACACACAAATCAATGTGCAAAATATCAAGCAATCTTAT 6119



KW Hypothetical protein: RNA-directed DNA polymerase.  
SQ SEQUENCE 1275 AA; 149043 MW; A93622F874A62CE4 CRC64;

## Alignment Scores:

Pred. No.: 0 Length: 1275  
Score: 5188.00 Matches: 1056  
Percent Similarity: 78.97% Conservative: 78  
Best Local Similarity: 73.54% Mismatches: 136  
Query Match: 9.91% Indels: 173  
DB: 4 Gaps: 7

US-10-083-853b-2 (1-29921) x Q9UN80 (1-1275)

```

QY 3428 ATGCGAGCATCAATTAATGACATATATATTAACCTTAATGTAATGGCTAAATTC 3487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MetThrGlySerThrSerHisIleThrIleuThrIleuAsnIleAsnGlyLeuAsnSer 20

QY 3488 CCAATTAAAGACACAGACTGGCAAAATGGATTAAGATCAAGACCATCAGTGTCTGT 3547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 AlaIleuysArgHisArgLeuAlaSerThrIleuysSerGlnAspProSerValCysCys 40

QY 3548 ATTGAGAGGCCCTTCACATGAAAGACACACATAGGCTCAAAATTAAGGATGGAGG 3607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 IleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTyrArg 60

QY 3608 AAGATTAAACCAAGTAATGGAACCAAAAAAAGAGAGGAGGTTGCAATCCTAGTCTCT 3667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 LysIleuYrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79

QY 3668 GATTAACACAGCTTTAAACCAACAGATCAAAAGACAAAGAGGCCATTACATAATG 3727
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 AspLysThrAspRheLysProThrLysIleLysArgAspLysGlnGlyHisTyrIleMet 99

QY 3728 GTAAGGATCAATGAAGACAGAGCTACTATCCAAATATACATGACCCCAATCA 3787
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ValLysGlySerIleGlnGlnGlnGlnLeuThrIleuAsnIleTyrAlaProAsnThr 119

QY 3788 GGAGCACCACAGATTCATTAAGCAAGTCTTAGAGACCTACAAAGACTTGGACTCCAC 3847
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139

QY 3848 ACAATAATAGTGGAGCTTAATAATTAATAGACATTTAACCCCACTCCCAATATTA 3907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 ThrIleuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153

QY 3908 GCGCATCAATGACACAGAAATTAACAGAGATATCCAGAGTTGAGCTGAGCTGGAC 3967
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnIleuAsnSerAlaLeuHis 173

QY 3968 CAAGCGACCTAATAGATATATACAGAACTCCCAACCCCAATCAACAGATATACATC 4027
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193

QY 4028 TTTCACAGCATCATACACCTATTTAAATTAAGTACCATGTAATTTTAAGTAAACACTC 4087
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213

QY 4088 CTCACGAAATGCAAAAGACGAATCTCTACCAACAGCTCTCTAGACTACAGTGCATC 4147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 LeuSerLysCysLysArgThrGlnIleThrAsnTyrLeuSerAspHisSerAlaIle 233

QY 4148 TATTTAAGCACTCAAAATTAAGAAACTCTCAAAATCTACCAACTACTACGAACTGAAAC 4207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 LysLeuGluLeuArgIleLysAsnLeuThrGlnIleSerArgSerThrThrLysLeuAsn 253

QY 4208 AACCTGCTCTGAATGACTACTGGTGAATTAATCAAAATGAAGCAAAATTAAGATGTC 4267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGlnIleuLysAlaGluIleLysMetPhe 273

QY 4268 TTGTAACCAATGACAAAGACACACATGACAGATCTCTGGGATTTTAAGCA 4327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 PheGluThrAsnGlnAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293

```

```

QY 4328 GTGTAGAGGGAATTTATAGCACTAGATGCTTCAAGAGAAAGCAAGAAATTTCAAA 4387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnLysArgLys 313

QY 4388 ATAGACACCTTAACATCAATTAATAAGAACTAGAGAAAGAAAGCAAAATTTCAAAA 4447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 IleAspThrLeuThrSerGlnLeuLysGlnLeuGlnLysGlnGlnThrHisSerLys 333

QY 4448 GCTAGCAGAAAGCAAGAAATTAATAGATCAGACAGACAGCAAGTGAAGATGACACAA 4507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 AlaSerArgArgGlnGlnIleThrLysIleArgAlaGluLeuLysGlnIleGluThrGln 353

QY 4508 AAGCCCTTCAAAATTAATCAATGAATCCAGAGAGCTGTTTAAAGAAATCAGCAAAAT 4567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 LysThrIleuGlnLys-IleAsnGlnLysArgSerTrpPheGlnArgIleAsnLysII 373

QY 4568 ----AGACCTACAGACATTAATTAAGAGAAAAGAGAGAAAGTAAGAGATGCAAT 4623
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 eAspArgProLeuAlaArgLeuIleLysLysArgGlnLysAsnGlnIleAspThrII 393

QY 4624 AAAAAATGATAAGGGGATATCACACCGATCCCAAGAAATTAACAATATTATACAGA 4683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 eLysAsnAspLysGlnLysPheThrThrAspProThrGlnIleGlnThrThrIleArgGI 413

QY 4684 ATATTATTAACACCTTATGCAATTAACCTAGAAATCTAGAGAAATGATTAATTCCT 4743
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 uTyrTyrLysHisLeuTyrAlaAsnLysLysGlnAsnLeuGlnIleuMetAspThrPheLe 433

QY 4744 GCACACATATGTACCTGTTATGACCTTGGGGACAGACAAAGGGGTAATGCAGAA 4803
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 uAsp----- 434

QY 4804 ATAAAGCAAAAGACAAAGAGTATGTTTGAAGTAGGGGTCAAGGGGCACTTGCCCT 4863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 4864 AATGACAGAGGCCCTGAGCTTACACACCCCTGTATTTATTAAGCAAAAGATAGC 4923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 4924 GAGAGGCTGAGTGGAGAGAGAGTCAAGCTGTTAGCTCCAGATAGCCCTGCAAGACTGC 4983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 4984 ATTCCCTCAAAATAGGCTAGATGTGCCAGTAGATTAACCTCAAGAGCCAGTGCAGG 5043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 5044 GAGTATGCCCCCTGACAAACCTTCTAGGGCAGGCACAGAGTAAGTTTGCCACATTC 5103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 5104 GTATTACAGATAAACAGTTGCTGTGTGATCAAGTAGCCCTCAAGTGCAGATTGG 5163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ----- 434

QY 5164 TCATGATGCCCTTTGGCCCTTTTGGCTCCCAAAACATACACCTCTCAAGACTTAACCA 5223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 ----- 435

QY 5224 GGAAGAGTCAAAATCCCTGAATATACAGATTAACAAGTTCTAAATTTGAAGCAGTAATGA 5283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 nGluGlnValGlnLeuSerLeuAsnArgProIleThrGlnIleSerGluIleValAlaIleIleAs 463

QY 5284 TAGCTTACCAACCAAAAAAAGATCCAGAGCAGACAGGATTCACAGCAAAATTTACAGAG 5343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrLysArgLysGlnArg 483

QY 5344 GTACAAAGAGAGGCTGACTATTCCTTGTCAAACTTTCACAAAATTAAGAA----AATGG 5400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 gTyrLysGlnGlnLeuValProPheLeuLeuLysLeuPheIleSerIleGluLysGlu-G 503

QY 5401 GAATCCCTCCCTAATCAATTTTACAGAGCCACATCAATCTGATACCAAAACTTAGAGTG 5460

```

Db 503 LylleuProaenSerPheTyrgLualaslelleleuileProlysProGlyArg 533  
QY 5461 ACACAAACAAAAGAGAAATTTTCAGGCCATTCCTGATGACATTTGATGTAATCC 5520  
Db 523 sphrThrlysluysgluasnPheargProileSerleuetaasnilleaspalylllel 543  
QY 5521 TCATAAATAATCTGGCAACCAATCCAGAGACATCAAAAGCTTATCTACATGATC 5580  
Db 543 euasnlyslleuallaasnArgillelInhlstleuylsluilehlshlsaprc 563  
QY 5581 AACTGGCGTCATCCCTGGATGTCAGGCTTGCTCAAAATATGCAATATCAATAGTAG 5640  
Db 563 lvalslPhelleleProGlymeGlnGlytrPhasnillearglylsSerlleasnVali 583  
QY 5641 GCCATCAATAAACACACCAATGACAAAAACACATGATTTATCTCAATGATGACGAAA 5700  
Db 583 leclInhlstleasnArglalyasplysasnhlshetlelleSerlleaspalyllle 603  
QY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAAAATTCAGTAAACTAGTATCG 5760  
Db 603 ysAlaPhaspplysilleGlnGlnProPhetleuylstrhleuasnlylsleuGlyllea 623  
QY 5761 ATNGATGTATCTCAAAATTAAGACTTATATAC-AAACCCACAGCCAAATATCATAC 5819  
Db 623 spelyThrlytrPheylsillelearglalletrysplysProthAlaasnillelel 643  
QY 5820 TGAATGGCAAAACGGAACATTCCTTTGAGATGGCACACAAAGATGCGCTC 5879  
Db 643 euasnlysluysleuGlnAlaPherProleuylstrGlythrArgGlnGlyCysProl 663  
QY 5880 TCTCACCACTCTATTCAAGATACTTGTGAAGTTCTGGCCAGGCAATCGGCAATAGA 5939  
Db 663 euserProleuLeuPhasnilleValleuGlnValleuAlaarglalleargGlnGlu 683  
QY 5940 AAGAATTAAGGTTATCAATAGAAAGAGAGAACTCATGCTGCTGTTGCGATG 5999  
Db 683 yselulleysgllyleGlnleuGlylsleuGlnVallylsleuSerleuPhasnlyspa 703  
QY 6000 ACATGTTTGTATATTAGAAAACCCATCGCTCAGGCCAAAACTCTTAAGCTGATTA 6059  
Db 703 spetilleVallyrleuGlnuasnProilleValserAlaglnasnleuLeuyllelles 723  
QY 6060 GCACCTTCAGCAAGCTCAGGACACAAAATCAATGTGCAAAAATACAAAGCTTTAT 6119  
Db 723 etasnPheserlyslValserGlytyrlyslleasnValGlnlylsSerGlnAlaPhelert 743  
QY 6120 ACCCAATATATAGACAAACAGAGAGCCAAATCATGAGTGAACCTCATTCACAAATGCTA 6179  
Db 743 yrThrlnasnasnarglnThrGlnSerGlnllemetGlyluLeuProPhetThrleleas 763  
QY 6180 CAAAGAGAAATAAATACCTAGGAATACAACTTACAAAGGACAGCTAGGAATCTTCAAG 6239  
Db 763 ertlystrGlyleuGlylleGlnleuThrArgaspVallylsaspleuPhelysg 783  
QY 6240 AGACACTCAACACATCATCAGAAATAAGAGAGACACAAACAAATGAGAAAACATTC 6299  
Db 783 llnasnlyrlyslProleuLeuylsluileuGlnlythrlnsnlystrlylsasnillep 803  
QY 6300 CATGCTCACAGATAGTAAGAAATCAT-----GAAATGCACTACGCCCAAGTAATTA 6352  
Db 803 rocyserTrpyalGlyArglleasnilleVallylsmeclAlalleuProlysValillet 823  
QY 6353 AATGATTCAGTGTCCCATCATCAAGCTTACATTTGACTTTCTTCACAGAAATGGAAGAAA 6412  
Db 823 yrAlrghasnAlaalleProilleuylsluLeuProkethrPhetethrGlnleuGlnlyst 843  
QY 6413 CAACCTTAATTTATGATGGAACCAAAAGAGCCACAGAGCCAGCAAGCAATCTTAAC 6472  
Db 843 hrThrleuylslPhelletrPasnGln-LysArgAlaArgilleAlaLysSerilleuSer 862  
QY 6473 AAAAGAAACAAAGCTGAGTATCATGCTGATTAACCTATACATTAAGGCTACA 6532

Db 863 GlnlylsasnlyslaleGlylylleThrleuProaspPhelylsleuTyrrlylsAlaThr 882  
QY 6533 GTAACCAAAACCTCATGTACTGTACCAAAACAGATATATAGACCAATGGAACAGAGA 6592  
Db 883 ValThrlyslThrAlaThrpyrrlyrGlnlnasnArgasplespGlntrPasnArglnr 902  
QY 6593 GAGACCTCAGAAATTAAC-ACGCAATCTACATCCATCTGATCTTTACAAACCTGACAAA 6651  
Db 903 GluProSerGlnlleuMetProhlslleTyAsnTyrrleuillePhasnlyslsProGlylys 922  
QY 6652 AACACAGCATGGAAGAAAGATTCCTTATTAATAATGGGTGGGAAAAAGCGTAGGC 6711  
Db 923 AsnlyslGlntrPlylsaspsSerleuPhasnlyslrPcystripolunsnrpleuAla 942  
QY 6712 ATATGCAAGAAAGCTGAACATGATGATCCCTTCCTTACACCTTATACAAAGTAACTCAGA 6771  
Db 943 lleCysArglylsleuylsleuasnProPhetleuThrProtyrThrlyslleasnserArg 962  
QY 6772 TGAATTAAGACTTAATATTAAGACATTAACCATTAACCA-CAAGAAAACCTAGGC 6830  
Db 963 TrpilleysaspLeuasnVallylsProlystrlleysThrleuGlnluasnleuGly 982  
QY 6831 AATACCATTCAGATATGGACATGGCAAGCTTCATGACTTAACACCAAGCAAGCAATG 6890  
Db 983 llerThrilleGlnspilleGlyalGlylsaspsPhemeSerlystrProlysAlaMet 1002  
QY 6891 GCACAAAGCCCAAAATAGACAAAGTGTGATTAACATATAGACTTGTGCACAGCA 6950  
Db 1003 AlaThrlylsaspplyslleasnlystrPasnpleuileylsluLeuylserPheCysThrAla 1022  
QY 6951 AAAAAAAGCTCATCAGAGTGAACAAGCAACCTACAGAAATGGGAGAAAATTTTGCAT 7010  
Db 1023 -LysGlnThrThrleargValasnArglnProThrTrpGlnlysllePheAlaTh 1042  
QY 7011 CTATGATGTGCAAGGCTTAATATCCAGAGCTCTGCAAGAACTTAACAAATTAACAA 7070  
Db 1042 rTyrrSerSerlyslGlyleuileSerArglylleArglnleuGlnleuGlnleuTyrr 1062  
QY 7071 GAAAAA-----AACACCCCGTCAAAATATGCGCAAGATATGACGACACTTGCAAA 7126  
Db 1062 sluslystrlnasnasnProilleyslystrPalaTalsasPmetlAsnArgHlsPheSerly 1082  
QY 7127 AGAAGACATTTATGCGCAACAAACATATGAAAAAACCCTATCATTCGTGCTG 7186  
Db 1082 sglnuaspilleTyrrAlaAlaLyslyshlsmetylsCysSerSerleuAlailear 1102  
QY 7187 AGAAATGCAAAACAAAACACAGTGAACACATCCATGATGATGATGATGATGATGATGAT 7246  
Db 1102 geluMetGlnlleysThrThrmetArglyrHlsleuThrProvalArgmetAlailel 1122  
QY 7247 TAAAAAGTCAGAAACACAAATGCTGAGAGGATGAGAAATAGAACACTTTTCCA 7306  
Db 1122 eluslylsSerGlylsasnasnArgCysTrpArgGlyCysGlyGlnilleGlyThrleuLeuHl 1142  
QY 7307 CTTGTGTGAGATGTAATTAATGTTCAACCATTTGAGAGACAGTGTGAGATTCCTTAA 7366  
Db 1142 scystrPrPasnPcyslylsleuValGlnProleuTrpLysSerValTrpArgPheLeuAr 1162  
QY 7367 GGATCTAGAACAGAAATATCATTTGACCAGCAATCCCATTAAGTATGATTAACCAAA 7426  
Db 1162 gaspleuGlnleuGlnilleProPhasnProAlalleProleuLeuGlylleTyrrPron 1182  
QY 7427 GGAATTAATAATCTTATTAATAAGACACATGCACACATATGTTTATGACGACTGAT 7486  
Db 1182 nGlnTyrrlylsSerCysCysTyrrlylsasphrCysThrArgmetPheilleAlaAlaLeuPh 1202  
QY 7487 CACATAGCAAAAGACTTGAACCAACCAAAATGTCATAGTATGATGATGATGATGATGAT 7546  
Db 1202 eThrilleAlaLyslnrTrpPasnGlnProlyscysProThrMetlleaspTrpilleysly 1222  
QY 7547 AACATGGCAATATCACCATCAATATCATGACAGCAATAAAG-GATGAGCTCATGTC 7605  
Db 1222 smettrPhtslleTyrrThrmeGlnlyTyrrAlaAlaileylsasnaspGluPhellese 1242

QY 7606 CTTTGCAGAGATATGATGAGCTGGAAACCATCATCTGCAGAACTAAACAGAGACA 7665  
|||||  
Db 1242 rrrvevalglYthrtiretyleugluThrllelleleuSerlyseuSerGlnGlu 1262  
QY 7666 GAAACCAACACACCATGTCTCTACTGTAACTGGAGCT 7705  
|||||  
Db 1262 nlYsthrlynsIsargllePheSerleuileGlyGlyAsn 1275  
RESULT 6  
O8TE30  
ID O8TE30 PRELIMINARY; PRT; 1275 AA.  
AC O8TE30;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical 149.0 kDa protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21668188; PubMed=11810275;  
RA Benjes S.M., Morris C.M.;  
RT "A full-length and potentially active LINE element is integrated  
RT polymorphically within the IGL locus in a genomically unstable region  
RT of chromosome 22.";  
RL Hum. Genet. 109:628-637(2001).  
DR EMBL; AF421375; AAL50637.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 1275 AA; 149010 MW; B327D9D50A581764 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 1275  
Score: 5186.00 Matches: 1056  
Percent Similarity: 78.83% Conservative: 76  
Best Local Similarity: 73.54% Mismatches: 138  
Query Match: 9.91% Indels: 173  
Gaps: 7  
US-10-083-853b-2 (1-29921) x O8TE30 (1-1275)  
QY 3428 ATGGCAGAGCAATTCACATTAATTAATTAATTAATTAATTAATTAATTAATTC 3487  
|||||  
Db 1 MetThrGlySerInsSerHisIleThrIleleuThrleuAsnIleAsnGlyLeuAsnSer 20  
QY 3488 CCAATTAAAGACACAGACTGGCAATTTGATTAAGATCAAGACCCATCAGTGTCTGT 3547  
|||||  
Db 21 AlalIlelysrghIsarglleuAlaSerTrpIlelySerGlnAspProSerValCysCys 40  
QY 3548 ATTCAGGAGGCCATCTCCACATGAAAGACACACATGCGTCAAAATTAAGGATGGAGC 3607  
|||||  
Db 41 IleGlnIuThrHisleuThrCysArgAspThrHisArgleuIlelySgIlyTrpArg 60  
QY 3608 AAGATTACCAAGTAATGAAAAACAAAAAAGAGAGGCTGCAATCTAGTCTCT 3667  
|||||  
Db 61 lYsIleIeYrGlnIlaAsnGlyLysGln---LysIleYsIlaGlyValAlaIleleuValSer 79  
QY 3668 GATTAACACAGACTTTAAACACAAAGATCAAAAGAGACAAAGAGCCATTACATATG 3727  
|||||  
Db 80 AspIlyThrAspIleIysProThrIlySIlleYsArgAspIlySgIlyHisIlyIleMet 99  
QY 3728 GTAAGGCATCATGGAACAAGAAAGCTAACTATCTTAATATACATGGACCAATACA 3787  
|||||  
Db 100 ValIlySerIleYrIleGlnGlnGlnIleuThrIleleuAsnIleYrAlaProAsThr 119  
QY 3788 GGAGCAACCCAGATTCAATTAAGCAAGTCTTAGAGACCTACAAAGAGACTTGACTCCAC 3847  
|||||  
Db 120 GAlYAlaProArpHeIleYsGlnValleuSerAspIleuIlnArgAspIleAspSerHis 139  
QY 3848 ACAATATAGTGGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3907  
|||||

Db 140 ThrleuIleMetGly-----AspPheAsnThrProleuSerThrleu 153  
QY 3908 GGCAGATCAATGAGACAGAAATTAACAGGATATCCAGAGTGAAGCTGAGCTGGAC 3967  
|||||  
Db 154 AspArgSerThrArgGlnIlyValAsnIlyAspThrGlnIleuAsnSerAlaIleuHis 173  
QY 3968 CAAGCGACCTAATAGATATCTACAGAACTCCCAACCCCAAACTCAACAGAAATATACATC 4027  
|||||  
Db 174 GlnAlaAspIleuIleAspIleYrArgThrleuHisProIlySerThrIuYrThrPhe 193  
QY 4028 TTCTCAGCATCATTACACCTATTTAAATTGACCATGTAAATTAAGTAAACATC 4087  
|||||  
Db 194 PheSerIaProHisIsthrItySerIySIlleAspHisIleValGlySerIySAlaIleu 213  
QY 4088 CTAGCAAAATGCAAAACAGAAAGAAATCCCTAAACAAACAGTCTCAGACTCAGTGAATC 4147  
|||||  
Db 214 LeuSerIySylsArgThrGluIleIleThrAsnIyIleuSerAspHisSerAlaIle 233  
QY 4148 TATTTAGACTCAGAAATTAAGAAATCTCATAAATCCACAACTACATGGAACCTGAC 4207  
|||||  
Db 234 LysleuGluIleuArgIleIySAsnleuThrGlnSerHisSerThrItrIlySleuAsn 253  
QY 4208 AACCTGCTCTGAATGACTACTGGGTAAATTAACAAATGAAGCAAAATTAAGAGTTC 4267  
|||||  
Db 254 AsnleuIleuAsnAspTyTrpValHisAsnGluMetIySAlaGluIleIySmetPhe 273  
QY 4268 TTGGAACCAATGGAACAAAGACACAAATGTACCAAGAAATCTCGGGCATATTTAAAGCA 4327  
|||||  
Db 274 PheGluThrAsnGlnAsnAspThrIthTyGlnAsnleuItrPaspAlaPheIySAla 293  
QY 4328 GTGTGTAGAGGAAATTTATAGCACTAGACTGCTTACAAAGAAAGCAAAATATCAAAA 4387  
|||||  
Db 294 ValICysArgGlyIySPhelIeAlaIleuAsnAlaItyrIySArgIySglnIuArgSerIyS 313  
QY 4388 ATAGACACCTTAATCATCAATTAATTAAGAACTAGACAGAAAGACAAACAAATTCAAA 4447  
|||||  
Db 314 IleAspThrleuThrSerGlnleuIySgIleuGlnIySglnGlnItrHisSerIyS 333  
QY 4448 GCTAGCAGAGACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4507  
|||||  
Db 334 AlaSerArgArgGlnGlnIleThrIySIlleArgIlaGluIleuIySgIuIleGluThrGln 353  
QY 4508 AAAGCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4567  
|||||  
Db 354 LysThrleuGlnIyS--IleAsnGluSerArgSerTrpPheGluArgIleAsnIySIl 373  
QY 4568 ----AGACACTAGACAGACTAATTAAGAAAGAAAGACAGAGATCAAGATGCAAT 4623  
|||||  
Db 373 eAspArgProleuAlaItrGleuIleIySlySArgGluIySAsnGlnIleAspThrIl 393  
QY 4624 AAAAATGATTAAGGGGATATCCACACCGATCCACAGAAATCAAACTTTATTCAGAGA 4683  
|||||  
Db 393 eIySAsnAspIySlySAspIleThrItrAspProThrGlnIleIlnThrIleArgI 413  
QY 4684 ATATTATTAACACCTCTATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTC 4743  
|||||  
Db 413 uTyTyTyIyGlnHisleuIyrgIlaAsnIySleuGlnIlaAsnleuGlnIuMetAspThrPhe 433  
QY 4744 GGACACATATGTAGCTGTATGACCTTGGGGGACAGAACAAAGGGGGTGAATGCGAA 4803  
|||||  
Db 433 uAsp----- 434  
QY 4804 ATTAAGACAAAGACAAAGAGATATGTTTGAAGTAGGGTCAAGGGGCAACTTCCTCT 4863  
|||||  
Db 434 ----- 434  
QY 4864 AATGACAAAGGCGCTGAGCTTTACACACCTCTGTATTTATTAAGCAAAAGATAGC 4923  
|||||  
Db 434 ----- 434  
QY 4924 GAGAGGGTGAAGGAGAGGAGTCACTGTTAGCTCCAGACTAGGCTGCAAGACTGC 4983  
|||||  
Db 434 ----- 434

```
OY 4984 ATTCTCAACAATAGGCTCTAGATGTCCAGTAGATTAACCTCAAGAGCCAGTGCAGG 5043
Db 434 ----- 434
OY 5044 GAGTGTAGGCCCTCAGCAAAACCTTCTAGAGGAGGACAGAAAGTAAGTTGCCACATTC 5103
Db 434 ----- 434
OY 5104 GTATTCACGATTAACAGATTGTGCTTTGATCAAGTAGCCTCCAGTGAATGCTGAGTTGG 5163
Db 434 ----- 434
OY 5164 TCATGATCCCTTTGGCTTTTGGCTCCAAAACATACACCCCTGTCAAGATAACCA 5223
Db 435 -----ThrTyrThrLeuProArgLeuasnG 443
OY 5224 GGAAGAAGTCAATTCCTGAATATACAGTAACAGTAATTCGAAATGAAAGCATATGCA 5283
Db 443 nGluGluValGluSerLeuasnArgProIleThrGlySerGluIleValAlaIleIleAs 463
OY 5284 TAGCCTACCAACCAAAAAAGTCCAGAGCAGACGAGATTCACAGCCAAATTCACAGAG 5343
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg 483
OY 5344 GTACAAAGAGAACCTGGTACTATCTCTTGAAGCTATTCCAAAAAATAGAA--AATGG 5400
Db 483 gTyrLysGluGluValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503
OY 5401 GATCCTCCCTAATCATATTTACAGAGCCAGCATCATCTGATACCAAAACCTACAGAGT 5460
Db 503 LysLeuProAsnSerPheTyrGluAlaSerIleLeuIleProLysProGlyArgA 523
OY 5461 ACACAACAAAGAGGAAATTTCCAGGCCCATATCCCTGATGAACATGTGTAATGTC 5520
Db 523 spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleI 543
OY 5521 TCATATAAATCTGGCAAAACCAATCCAGAGCAGATCAAAAAGCTTATCTACCATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleAspG 563
OY 5581 AAGTTGGCGTATCCCTGGAGATGCAAGGCTGGTCAAAATGCAAAATGCAAAATGTA 5640
Db 563 lnaAlaLysPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 583
OY 5641 GCCATCACATAACAGAACCAATGACAAAACACATGATTTATCTCAATAGATGACAAA 5700
Db 583 lGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGlu 603
OY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATCTCACTAAACTAGTATCG 5760
Db 603 ysaAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623
OY 5761 ATGGATGTATCTCAAAATTAATAGAGCTATTTATAC-AAACCCACAGGCATATATAC 5819
Db 623 spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIle 643
OY 5820 TGAATGGCAAAACGAGAGCATTCCTTTGAGACTGGCACAAAGACAGAGATGCCCTC 5879
Db 643 euAsnGlyLysIleGlnLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 663
OY 5880 TCTCACCACTCTATTCAAGATACTATTTGGAAGTTGCGCCAGGCAATCAGGCATATAG 5939
Db 663 euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGlu 683
OY 5940 AAGAAATAAAGGCTATCAAAATAGAAAGAGAGAAATCATATTGTCTGTTTGCAGATG 5999
Db 683 ysaGluLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 703
OY 6000 ACATGTTGTATATTAGAAAACCCATGCTGTACAGGCCAAAACCTCTTAGCTGATTA 6059
Db 703 spMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIle 723
OY 6060 GCAACTTCAGCAAAAGTCTCAGAGACACAAAATCAANTGCAAAAAATCAGACATTCATT 6119
Db 723 eraSnPheSerLysValSerGlyTyrIleAsnValGlnLysSerGlnAlaPheLeu 743
OY 6120 ACCCAATTAATAGCAAAACAGAGCCAAATCATATGATCACTCATCAATAATGCA 6179
Db 743 yThrAsnAsnArgGlnThrGluSerGlnIleMetLysIleLeuProPheThrIleAla 763
OY 6180 CAAGAAGAAATTAATACCTAGGAATCAACTTACAGAGGACAGTGAAGACTCTTCAAG 6239
Db 763 erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspAlaLysAspLeuPheLysG 783
OY 6240 AGAATACAAACCACTGATCAAGAAATAGAGAGACACAAACAAATGAAAACATTC 6299
Db 783 lAsnTyrLysProLeuLeuLysGluIleLysGlnLysPheAsnLysTrpLysAsnIleP 803
OY 6300 CATGCTCACAGATAGTAAAGATCAT-----GAAAATGCCATCTAGCCCAAGTAAAT 6352
Db 803 rocYserTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
OY 6353 ATAGATTCAAGTGTACCCCATCAAGCTACATTCATCTTCCACAGAAATGAAAAA 6412
Db 823 yArgPheAsnAlaIleProIleLysLeuProMetThrPheThrGluLeuGluLysT 843
OY 6413 CAACTTAATTTATATATGAAACCAAAAAAGACCCACAGAGCCACAGCAATCTTACG 6472
Db 843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaIleArgIleAlaLysSerIleLeuSer 862
OY 6473 AAAAAGAACAAAGCTGGAGTATCATGCTACCTGACTTAAACTATATTAAGGTACA 6532
Db 863 GlnLysAsnLysAlaIleGlyIleThrLeuProAspPheLysLeuTyrLysAlaThr 882
OY 6533 GTAACCAAAACGTCATGTACTGTGTACCAAAACAGATATATAGACCAATGGAACAGACA 6592
Db 883 ValThrLysThrAlaTrpTyrTrpGlnAsnArgAspIleAspGlnTrpAsnArgThr 902
OY 6593 GAGACCTCGAATATAC-ACGCAATCTACATTCATCTGATCTTTGACAAACCTGCACAA 6651
Db 903 GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys 922
OY 6652 AACAAAGCAATGGAAGAAAGTCCCATTTAATTAATGTTGGAAGAAACGCGTAGCC 6711
Db 923 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
OY 6712 ATATGCAAGAAACGTAACGATCCCTTCCTTACACCTTATACAAAGTAACTACA 6771
Db 943 lLecysThrLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
OY 6772 TGAATTAAGACTTAATATTAAGACATTAACCATTAACCA-GAAGAAAACCTAGCC 6830
Db 963 TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGlnGluLysLeuGly 982
OY 6831 AATACCATTCAGATATGACATGGGCAAGAGCTCATGATTAACCAACCAAAACCAATG 6890
Db 983 lIleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
OY 6891 GCAACAAAAGCCAAATAGACAAAGTGGATCTGATTTAACTATAGACTTCGACAGCA 6950
Db 1003 AlaThrLysValLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
OY 6951 AAAAAAACTGTATCAGAGTGAACAGACACCTACAGAAATGGAGAAAATTTTGCAAT 7010
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaThr 1042
OY 7011 CTATGATCTGCAAAAGGCTAATATCCAGAGATCTACAGAAAGCAATTAACAAATTTACA 7070
Db 1042 rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLys 1062
OY 7071 GAAAAA---AACACCCCGCTCAAAATATAGGCAAGATATAGACAGCACTTCTCAAA 7126
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys 1082
OY 7127 AGAAGACATTTATGACGACCAACAAACATATGAAAAAAACCTCATCATGCTTGCTTAG 7186
```



```

Db      1082 sgluasprlethylalaalalsylshismetlylscyserserleuallalear 1102
QY      7187 AGAATGCAAAACAAACACAGTGCATACATCTATGCTAGTATGATGGATGAC 7246
Db      1102 gglmetelalethylshirhmetlarythshislenlthprovalargmetalaileil 1122
QY      7247 TAAAGAGTCAGAAACAAACAAATGCTGAGAGAGATGTGAGAAATAGAACACTTTTCCA 7306
Db      1122 elyslysserelysnsmnargcyltrparagelycysglyolulegylthrlleuunh 1142
QY      7307 CTGTGGTGGAGATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7366
Db      1142 scyltrprpaspysylsleuvalglinproleuthrlpyservaltrparphleuar 1162
QY      7367 GGATCTGAGAACCAAAATATATATATGACCCGACCAATCCCATTTACTGAGTATATCCCAA 7426
Db      1162 gaspelengluleuulileprophasprioalaleproleuengllyleryproas 1182
QY      7427 GAATATTAATCATCTATTAATTAAGACACATGACATCATGTTTATTCAGACACATGAT 7486
Db      1182 ngultrlylsercyscyltrlylaspthrlylaspthrlylaspthrlylaspthrlylaspthr 1202
QY      7487 CACAATAGCAAGACCTTGGAACCAACCAATGTCATCAGTATGATGATGATGATGATGATGAT 7546
Db      1202 ethrllealalyshlrtrpansnglnleuylscysprothrmetileasprlileulsly 1222
QY      7547 AACATGGCACATATATACACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7605
Db      1222 smettrhishislerlythrlmetglulyltralaalalelysasnsapcgluphilese 1242
QY      7606 CTTCGACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665
Db      1242 rphelvalgllythrltrpethllyslleugluthrlleileuenserlylseuerglnglul 1262
QY      7666 GAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 7705
Db      1262 nlysthrlyshislarlylpheserleuileglyllyasn 1275

RESULT 7
ID      000366 PRELIMINARY; PRT; 1275 AA.
AC      000366
DT      01-JUL-1997 (TREMBLrel. 04, Created)
DT      01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Putative p150.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97285120; Pubmed=9140393;
RA      Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA      Debernardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT      "Many human LI elements are capable of retrotransposition."
RL      Nat. Genet. 16:37-43(1997).
DR      EMBL; U93567; AAC51267.1; -.
DR      InterPro; IPR005135; Exo_endo_phos;
DR      InterPro; IPR00477; RVTse.
DR      Pfam; PF00372; Exo_endo_phos; 1.
DR      Pfam; PF00078; rvt; 1.
KW      RNA-directed DNA polymerase.
SQ      SEQUENCE 1275 AA; 149033 MW; 07E88F8F4DB831A2 CRC64;

```

```

US-10-083-853B-2 (1-29921) x 000366 (1-1275)
QY      3428 ATGCGAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3487
Db      1 Metthrlyserasnsenrshislerlthrlleuuthrlleuashnleasngllyleuasner 20
QY      3488 CCAATTAAGACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 3547
Db      21 Alalleysarshislarlyleualaserlthrlleuenserllyleuaserlthrlleuaser 40
QY      3548 ATTGAGAGGCCATCTACATGAAAGACACACACACACACACACACACACACACACACACACACACAC 3607
Db      41 lileglnluthrshislerlthrcysargaspthrshislarlyleuylslyltylparg 60
QY      3608 AACATTTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3667
Db      61 Lysllyltylrglnlaasnngllylsgln--Lysllyllyllyllyllyllyllyllyllyllyllyl 79
QY      3668 GATAAAGACACCTTAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3727
Db      80 AspLysThrAspThrLysThrLysThrLysThrLysThrLysThrLysThrLysThrLysThrLysThr 99
QY      3728 GTAAAGCATCATGAAACCAAGACACACACACACACACACACACACACACACACACACACACACAC 3787
Db      100 Vallylsglyserlileglnlglnglulglulglulglulglulglulglulglulglulglulglul 119
QY      3788 GGAGACCCACATTCATTAAGCAAGATCTTATGACACCTACACACACACACACACACACACACACAC 3847
Db      120 Glylalarproargprrhilellylsglnvalleuenseraspeluglnargaspelubaspelub 139
QY      3848 ACATATTAAGAGGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3907
Db      140 Thrleuilemetgly-----AspPheasnThrProleuSerThrleu 153
QY      3908 GCGACATCAATGAGACACAAATTAACAGAGATATCCAGAGATGGAAGTGAAGTGAAGTGAAGTGAAG 3967
Db      154 Hisargserthrarglnllylvalasnlylaspthrnglulglulglulglulglulglulglulglul 173
QY      3968 CAACGGACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4027
Db      174 Glinalaspelulleasprilellylarylthrlleuaspelubaspelubaspelubaspelubaspelub 193
QY      4028 TTTCACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4087
Db      194 Pheaserlarprrhishislerlthrllylserlyllylshislerllyllyllyllyllyllyllyllyl 213
QY      4088 CTCAGCAATGCAAAAGACAGAAATCTTAACAAACAGCTCTCAGACCTACAGCTACAGCTACAGCTAC 4147
Db      214 Leuenserlyscylsarglthnglulilelthrsnlyllyllyllyllyllyllyllyllyllyllyllyl 233
QY      4148 TATTTAGACTCAGAAATTAAGAACTCAGCTCAACAAATCAACACACATCAATGGAATGGAATGGAAT 4207
Db      234 Lysleuengluleuargllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 253
QY      4208 AACCTGCTCCGATGACTCTGCTGTAATTAACAAATGAAGCAAAATTAACAAATTAACAAATTAACAAAT 4267
Db      254 Asnleuenuleuasnsplyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 273
QY      4268 TTTGAAACCAATGAGAACCAAGACACACATGACCAAGATCTCTGGGCAATTAATTAAGCA 4327
Db      274 Pheglulthrnsnglulnlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 293
QY      4328 GTGTGTAGAGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4387
Db      294 Vallyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 313
QY      4388 ATAGACACCTTAATCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4447
Db      314 lileasprhrlleuuthrsergluleuylsglulglulglulglulglulglulglulglulglulglulglul 333
QY      4448 GCTAGCAGAAACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4507

```



[illegible]

QY	5581	AAGTTGGCGTCATCCCTGGGATGCAAGCGCTGGTTCAAAATATGCAAAATCAATAAATGAG	5644
Db	563	InvalidPhelIeProGlymetGInGlyTTPheAsnIleArgIysSerIleAsnValI	583
QY	5641	GCAATCACATAAACAGAACCAATGACAAAACACATATATATTCATATAGTCAGCAAA	5700
Db	563	IeGInHisIleAsnArgAlaIyAsnAspIySAsnHISmeIleIleSerIleAsnArgIuL	603
QY	5701	AGGCCCTTGTCAAAATTCACAGCCCTTCATGCTAAAAATTCAGTAGTAAGTATG	5760
Db	603	ysAlaPheAspIySIlleGInGInProPheMetLeuIyThIleuAsnIySleuGlyIleA	623
QY	5761	ATGCAATATATCTCAAAATTAATTAAGAGCTATTATTAAC-AAACCCACACCCCAATATCAAC	5819
Db	623	SPGIyThIyTTPheIySIlleIleAlaGAlaIleTyAspIySProThrIleAsnIleIleI	643
QY	5820	TGAATGGGCAAAAACCTGGACACATTCCTTTGAGAACCGGACCAAGACAGACATGCGCTC	5879
Db	643	euaSnGlyGInIySleuGInuAlaPheProIeuIySThGlyThIArgIInGlySProL	663
QY	5880	TCCTACACCTCCTATTCAGATACTATTGGAAGTTCTGGCCAGGGCAATCAGGCANTGA	5939
Db	663	euserProIeuIeuPheAsnIleValIleuGInuValIleuAlaArgAlaIleArgGInGInuL	683
QY	5940	AAGAAATTAAGGATTCAAATTAGAAGAAGAGGAAGATATATTCCTGTTGAGATG	5999
Db	683	ysGInIleIySgIyIleGInIeuGInIySgInGInuValIyIySleuSerIeuPheAlaAspA	703
QY	6000	ACATGTTTGATATTTAGAAAAACCCCATCGTCTCAGGCCAAAAACCTCTTAAGCTGATA	6059
Db	703	spmeIleValIyIleuGInuSnpProIleValSerAlaGInAsnIleuIySleuIleS	723
QY	6060	GCACTTCAGAAAGCTCTCAGACACAAATATGTCACAAATATTCACAACTTCAT	6119
Db	723	erasnPheserIySAlaSerIyTyIyIySIlleAsnValGInIySserIInIaPheIeuT	743
QY	6120	ACGCCAAATTAAGACAAACAGAGAGCCAAATCATAGTAGAATCTTCATTCAAATTGCTA	6179
Db	743	yThIAsnAsnArgIInThrGInuSerGInIleMetGlyGInIeuProPheThrIleAlas	763
QY	6180	CAAGAGACATTAATTAATCTAGGAATACAACTTACAGAGGACACAGTAGAACTCTTCAAG	6239
Db	763	erIySArgIleIySArgIleuGInIyIleGInIeuThrIArgAspValIySAspIeuPheIySg	783
QY	6240	AGAACTACAAACCACTGATCAAGGAATAATTAAGAGAGACACAAACAAATGCAAAACATTC	6299
Db	783	IuAsnTyIySProIeuIeuIySgIuIleIySgIuAspThrIAsnAsnIySAsnIleP	803
QY	6300	CATGCTCACAGATAGTAAGATCAT-----GAAATGCCATCTGCGCCAAAGTAATT	6352
Db	803	roCysSerIyValGlyArgIleAsnIleValIySMeAlaIleIeuProIySValIleT	823
QY	6353	ATGATATTCAGTATCCGCCCATTCAGGATACATGACTTCTTTCACGAATTTGGAAAAA	6412
Db	823	yArgPheAsnAlaIleIeuProIleIySleuPProMetThrIPherPheThGInIeuGInuIyST	843
QY	6413	CAACTTTAAATTCATATGAGACCAAAAAAAGACCACAGAGCCACAGACATTTAAGC	6472
Db	843	hThIleuIySPhelIerIyPAsnGIn-LyAsrArgAlaArgIleAlaIySserIleIeuSer	862
QY	6473	AAAAAGAACAAAGGTGAGGTATGATGCTACCTGACTTAAACATATCTATTAAGGCTACA	6532
Db	863	GInIySAsnIySAlaGlyGlyIleThIleuProIAspPheIySleuTyTyTyIySAlaThr	882
QY	6533	GTAAACAAAATCGATGGTACTGTTACCAAAACAGATATATAGACCAATGGACAGAACAA	6592
Db	883	ValIhIySThrIaIeATpTyTTPTyGInAsnArgIAspIleAsnGInIyPAsnArgThr	902
QY	6593	GAGACCTCAGAAATTAC-ACTCAGATTCATCATGTGATCTTTGACAAACCTGCACAA	6651
Db	903	GInuProSerGInIleMetProHisIleTyAsnTyIleuIleIeuPheAspIySProGlyIyS	922

```
QY 6652 AACAGCAATGAAAAAGATTCCATATTTAAATGCTGTGGAAAAACGTCAGCC 6711
    |||
Db 923 AsnLysGlnTrpClyLysAspSerLeuPheAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
QY 6712 ATATGCAAGAAAGCTGAACTGGATCCCTTCCTTACACCTTATACAAAAGTTAACTCAAGA 6771
    |||
Db 943 ValCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATATACACATAAACCATAAAAACCA-GAAGAAAAACCTAGGC 6830
    |||
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnAsnLeuGly 982
QY 6831 AATACCATTCAGATATAGCATGGCAAGCAAGCTCATGATGATAAACACCAAAAGCAATG 6890
    |||
Db 983 IleThrIleGlnAspIleGlyValIleLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCACAAAAAGCCAAATATACAAAGTGGATCTGATTAACTATAGAGCTTCACAGCA 6950
    |||
Db 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAAGCTCATCAGAGTGAAGCAAGCAAGCTACAGATGGGAGAAATTTTGCAT 7010
    |||
Db 1023 -LysGlnThrThrIleArgValAsnArgGlnProThrThrTrpGlnLysIlePheAlaThr 1042
QY 7011 CTATCGATCTGACAAAGGCTAATATCCAGATCTACGAAAGACTTAACAAATTTACAA 7070
    |||
Db 1042 rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyrLys 1062
QY 7071 GAAAAA----AACACCCCGCTCAAAATATGGGCAAAAGATATAGCAGACACTTCTCAA 7126
    |||
Db 1062 sLysLysThrAsnAsnProIleLysTrpAlaLysAspMetAsnArgHisPheSerLys 1082
QY 7127 AGAAGACATTTATGCGCCCAACATATGAAAAAAGCTCATCATGCTGCTGCTGCTGCTG 7186
    |||
Db 1082 sGlnAspIleTyrAlaLysLysHisMetLysLysCysSerSerLeuAlaIleArg 1102
QY 7187 AGAATGCAAAACCAACACAGCTGATCATCATCATGCTGATGATGATGATGATGATGAT 7246
    |||
Db 1102 gGlnMetGlnIleLysThrThrMetLArgTyrHisLeuThrProValArgMetAlaIleIle 1122
QY 7247 TAAAAAGTCAGGAAAAACAATGCTGGAGAGATGTGGAAAAATAGGAACACTTTTCCA 7306
    |||
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGlnIleLeuIleThrLeuLeuH 1142
QY 7307 CTGTGTGGGGAATGTAATTTAGTTCACCATGTGGAAAGACGTGGGAGATTCTCTAA 7366
    |||
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 1162
QY 7367 GGATCTAGAACCCAGAAATATATCATTTGACCAGCAATCCCATTAAGTATATACCCAAA 7426
    |||
Db 1162 gAspLeuGlnLeuGlnIleProPheAspProAlaIleProLeuLeuGlyIleTyrProLys 1182
QY 7427 GGAATTAATATCTTCTATTATTAAGACACATGCACACATATGTTTATTCACACTGAT 7486
    |||
Db 1182 sAspTyrLysSerCysCysTyrLysTrpHisTrpHisArgMetPheIleAlaLeuPhe 1202
QY 7487 CACAATAAGCAAAAGCTTGAACCAACCAATGTCATCATGTCAGTACAGCTGATAAGAA 7546
    |||
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleLysTrpIleLysLys 1222
QY 7547 AACATGGCATATATACACATGAAATATATGATGAGCCATTAAG--GATGAGTTCATGTC 7605
    |||
Db 1222 sMetTrpHisIleTyrThrMetGlnTyrTrpAlaAlaIleLysAsnAspIleuPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGAGTGGAAACCATCATTTCTAGCAAACTAACAAGACACA 7665
    |||
Db 1242 rPheValIleTyrThrMetLysLeuGlnThrIleIleLeuSerLysLeuSerGlnGln 1262
QY 7666 GAAACCAACACACATGTTCTACATTAAGTGGAGAT 7705
    |||
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275
RESULT 8
```

```
000375
ID 000375 PRELIMINARY; PRT; 1275 AA.
AC 000375;
DT 01-JUL-1997 (T-REMBLrel. 04, Created)
DT 01-JUL-1997 (T-REMBLrel. 04, Last sequence update)
DE 01-JUN-2002 (T-REMBLrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; Pubmed=9140393;
RA Sassenan D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Neas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93572; AAC51276.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVase.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvc; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 14905 MW; 456DD0F3DD7A17F5 C6C64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5182.00 Matches: 1054
Percent Similarity: 79.04% Conservative: 81
Best Local Similarity: 73.40% Mismatches: 135
Query Match: 9.90% Indels: 173
DB: Gaps: 7

US-10-083-853b-2 (1-29921) x 000375 (1-1275)
QY 3428 ATGCGAGATCAATTCACACATATATATTAACCTTAATTAATGAGGCTAATATCC 3487
    |||
Db 1 MethThGlySerAsnSerHisIleThrIleValThrLeuAsnGlnLysLeuAsnSer 20
QY 3488 CCAATTTAAAGACACACAGACTGGCAAAATGGATTAAGATGACAGACCATGATGCTGT 3547
    |||
Db 21 AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTCAGAGGCCCATCTGCACATGAAAAAGACACACATAGGCTCAAAATTAAGGATGAGG 3607
    |||
Db 41 IleGlnGlnThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTyrArg 60
QY 3608 AAGATTACCAAGTAATGGAACAAACAAAAAAGCAAGGGGTTCATCTAGTCTCT 3667
    |||
Db 61 LysIleTyrGlnAlaAsnGlnLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATTAACACAGACTTTAAACCAACAAAGATCAAAAGACAAAGAGCCATTACATPANG 3727
    |||
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGlnGlyHisTyrIleMet 99
QY 3728 GTAAGGCAATCATGGAACAAGAAGCTATCATCTTAATTTATCATGACCAATATCA 3787
    |||
Db 100 ValLysLysSerIleGlnGlnGlnGlnGlnLeuThrIleLeuAsnIleTyrAlaProAsnThr 119
QY 3788 GGAGCAAGCCAGATTCATTAAGCAAGTCTTAGAGACTCAACAAAGACACTTGACCTCCAC 3847
    |||
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATATAGTGGGAGCTTAATATATATATATAGACACTTTAACAACCCCACTGCCAATATTA 3907
    |||
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3908 GGAGATGCATGACAGAGAAATTAACAAAGATATACAGAGTGAAGTGAAGTGAAGTGAAG 3967
    |||
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGlnLeuLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGACCTAATATATCTACAGAACTCCCAACCCCAAAATCAACAGAAATATACACTC 4027
```

|||||  
Db 174 GlnAlaAspLeuIleAspIleTyrAlaThrLeuHisProLysSerThrGluTyrIlePhe 193  
OY 4028 TTCTCAGCATCACAATPACASCTATTTTAAATGACCATGATTAATTTTAATAAACATC 4087  
Db 194 PheSerAlaProHisIleThrIleTyrSerLysIleAspHisIleValIleLysSerLysAlaLeu 213  
OY 4088 CTCAGCAATGCAAAAAGAACAGAAATCTAACAAACAGTCTCTCAGATACAGTCGAATC 4147  
Db 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233  
OY 4148 TATTTGACCTCAGATTTAGAAACCTCACAATCACAACATCTCATGGAAACCTGAC 4207  
Db 234 LysLeuGluLeuArgIleLysAsnLeuSerGlnSerArgSerThrThrIlePheLysLeuAsn 253  
OY 4208 AACCTGCTCTGATGACTACTGGGTAATTAACAAATAGAGCAAAATTAAGATGTTTC 4267  
Db 254 AsnLeuTyrLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe 273  
OY 4268 TTTGAAACCAATGAGAACAAAGACACAATGTACCAATCTCTGGGCAATATTTAAGCA 4327  
Db 274 PheGluThrAsnGluLysAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293  
OY 4328 GTGTGAGAGGGAATTTATAGACATAGATGCTTACAGAGAAACGAGAAATATCTAA 4387  
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnLysArgSerLys 313  
OY 4388 ATAGACACCTTAAACATCACAATTTAAAGAACTAGAGAAAGAGCAAAACAAATTCAAA 4447  
Db 314 IleAspThrLeuThrSerIleLeuLysGluLeuGluLysGlnGluIleThrHisSerLys 333  
OY 4448 GCTACGACAGACAGACAAATTAACATTAAGTCAGACGAACTGAGAGAGATAGACACAA 4507  
Db 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353  
OY 4508 AAAGCCCTCAATTAATGAATGAATGCAAGAGCTGTTTGTGAAAGATCAGCAAAAT 4567  
Db 354 LysThrLeuGlnLys-MetAsnGluSerIleArgSerTrpPhePheGluArgIleAsnLysIle 373  
OY 4568 AGACCA----CTAGACAGACTAATAAGAGAAAGAGAGAGAAATCAAGATGCAAT 4623  
Db 373 AspArgLeuLeuAlaArgLeuIleLysLysLysArgLysAsnGlnIleAspThrIle 393  
OY 4624 AAAAATGATAAGGGGATATCACACCGATCCACAGAAATACAAATATTATTCAGAGA 4683  
Db 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGlu 413  
OY 4684 ATATTATAACACCTATGCAAAATTAAGTAAATCTAGAAAGAAATGATAAATTCCT 4743  
Db 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe 433  
OY 4744 GGACACATATAGCCCTGATGAGACCTGGGGGACAGAAACAAAGGGGGTGAATGCAGAA 4803  
Db 433 uAsp----- 434  
OY 4804 ATAAAGACAAAGACAAAAGATATGTTTGAAAGTAGGGGTACAGGGGCAACTTGCTCT 4863  
Db 434 ----- 434  
OY 4864 AATGACAAAGGCGCTGAGCTTTACACACCGCTGTATTATTAGGCAAAAGAGATAGC 4923  
Db 434 ----- 434  
OY 4924 GAGAGGTGAGTTGGAAGAAGAGTCAAGCTGTAGTCCAGATAGGCGCTGCAAGACTGC 4983  
Db 434 ----- 434  
OY 4984 ATTCTCAAAACAATAGGCTCTAGATGTCCCACTAGATACTCAAGAGACCGCAGTGCAGG 5043  
Db 434 ----- 434  
OY 5044 GAGTATGGCCCTCAGCAAAACCTCTAGAGGCAAGGACAGAAAGTAAGTTGCCACATCTCT 5103

Db 434 ----- 434  
OY 5104 GATATCCAGATAAACAGATTGGCTGTGATCAAGTAGACCTCCAGTGAATGCTGATTCG 5163  
Db 434 ----- 434  
OY 5164 TCATGATCCCTTTGGCCTTTTGGCTCCAAAACATACACCTCCCTCAAGCTAAACCA 5223  
Db 435 -----ThrTyrThrLeuProArgLeuAsnGlu 443  
OY 5224 GGAAGAAGCAATCCCGTGAATATACCAATCAAGTAAATGAAAGAGTAAATGTA 5283  
Db 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleLeuAs 463  
OY 5284 TAGCCTACCAACCAAAAAAGTCCAGACAGACGAGATTCCAGCCAAATTTACACAGAG 5343  
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg 483  
OY 5344 GTACAAAGAGAGCTGTACTTCTCTGAAACTATTCCAAAATATGAA----AATCG 5400  
Db 483 gTyrLysGluGluLeuValProPheLeuLysLeuPheGlnSerIleGluLysGlu-G 503  
OY 5401 GAATCCCTCACTCATTTTACGAGGCCAGATCATCTGATACCAAAACCTAGACAGT 5460  
Db 503 LysLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArg 523  
OY 5461 ACACAACAAAAGAGAAATTTACAGGCCATATCCCTGATCAACTGATGTGAAATTC 5520  
Db 523 sPThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 543  
OY 5521 TCAATTAATTAATGCGCAAAATCCAGACACATCAAAAAGCTTATCTACATGATC 5580  
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspArg 563  
OY 5581 AAGTGGCGTCATCCCTGGATGTCAGAGCGTGGTCAAAATGCAAAATCAATTAATGTG 5640  
Db 563 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleHisLysSerIleAsnValI 583  
OY 5641 GCCATCACATAAACAGAACCAATGACAAAAACCATGATTTATCTCATAGTGCAGAAA 5700  
Db 583 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGlu 603  
OY 5701 AGCGCTTTGTCAAATTTCAACAGCCCTTCATCTGTAATTTCTCAGTAACATGATTCG 5760  
Db 603 yAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyLea 623  
OY 5761 ATGGAATGATCTCAAAATTAAGAGCATTTATAC--AAACCCAGACCAATATCATAC 5819  
Db 623 sPGLyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIle 643  
OY 5820 TGAATGGGCAAAAACCTGGAAGCATTCCTTTGAGAACTGGCAAGACAGAGATCCCTC 5879  
Db 643 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysPro 663  
OY 5880 TTCTACACACTCTATTCAGAGATACATTTGGAAGTTCTGGCCAGGCAATCAGCAATAGA 5939  
Db 663 euSerProLeuLeuPheAsnIleValIleGluValLeuAlaArgAlaIleArgGlnGlu 683  
OY 5940 AAGAATTAAGGGTATTCAAAATAGAAAGAGAGAGATCATATTTCTGTTGAGAGATG 5999  
Db 683 ySGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAsp 703  
OY 6000 ACATGTTGTATATTAGAAAACCCATCGTCTCAGGCCAAAACCTCTTAAGCTGATPA 6059  
Db 703 sPmetIleValLysLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIle 723  
OY 6060 GCAACTTCAGCAAAATCTCAGACACAAATTAATGTCGCAAAAATCACAAGATCTTAT 6119  
Db 723 eAsnAspSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT 743  
OY 6120 ACGCAATATATGACAAACAGACAGCAAAATCATAGTACATCATCAATTTGCTTA 6179  
Db 743 yThrAsnAsnArgGlnThrGlnSerGlnIleMetSerGluLeuProPheThrIleAlaS 763

```

QY 6180 CAAGAGAAATAAATACCTAGAGATCAACTTACAAAGGACACAGTAGAAGCTCTCAAG 6239
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 763 eTysnrtgTlelystyTleuglyLleGlnleuthrtrgaspValysAspLeuphelysc 763
QY 6240 AGAAGTCAACAACACATGATCAGGAATAAGAGAGACACAAACAAATGGAAAAACATTC 6299
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 783 lAsnntgTlyspProleuLleuglyLleGlyslAspThrAsnlystrPlyAsnillep 803
QY 6300 CATGCTCACAGATAGTAAGATCAT-----GAAATGCCATACCTGCCAAAGTAAT 6352
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 803 rOcysserTrpValGlyArglleAsnilleValysMetAlalleleProLysValilleT 823
QY 6353 ATAGATTCAGTGTACCCCATCAGATCAGTACATGCTTCTTCACAGATTTGAAAAAA 6412
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 823 yTrArgPheAsnAlalleProLleLysleupProkethrPhepethrGlnleuglyust 843
QY 6413 CAACCTTAATTCATATGGAACCAAAAAAGAGCCACAGAGCCCAAGACATCTTAAGC 6472
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 843 hrThrleuLysPheilleTrpAsnGln-LysArgAlaArglleAlaLysSerilleLysSer 862
QY 6473 AAAAAGAACAAAGCTGGAGTATCATGCTACCTGACTTAATAACTATATCTTAAGGCTACA 6532
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 863 GlnLysAsnLysAlaGlyGlylleThrLeuproAspPheLysleuTrpHisLysAlaThr 882
QY 6533 GTAAACCAAACTGATGCTAGTCTGACCAAAACAGATATATAGACCAATGCAAGACACA 6592
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 883 ValThrLysThrAlaTrpTrpTrpTrpGlnAsnArgAspIleAspGlnTrpAsnArgThr 902
QY 6593 GAGACCTCAGAAATAC-ACAGCAATCTCATCTGATCTTGTGCAAACTGACAAA 6651
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 903 GluProSerGlnilleMetProHisIleTrpAsnTrpLeuIlePheAspLysProGlnLys 922
QY 6652 AACAAACAAATGAAAAAGATTCCTTATTAATTAATGGGTGGAAAACTGGCTAGCC 6711
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 923 AsnLysGlnTrpLysAspSerLeupheAsnLysTrpCysTrpGlnAsnTrpLeuAla 942
QY 6712 ATATGAGAAAGCTGAAAGTGCATCCCTTACACCTATATCAAAAGTAATCTCAGA 6771
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 943 IleCysArgLysleuLysLeuAspProPheLeuThrProTrpThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAATATATAGACATTAACCATAAACCA-GAAGAAAACTTAGGC 6830
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrleuGlnGlnAsnleugly 982
QY 6831 AATACCATTCAGATATGACATGGCAAGACTTCTATGCTTAACACCAAAAGCAATG 6880
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAAGCCAAAATAGACAAGTGGATCTGATTAACCTATATAGACCTTCGACAGCA 6950
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1003 AlThrLysAlaLysIleAspLysTrpAspLeuIleLysleuLysSerPheCysThrAla 1022
QY 6951 AAAAAAACTGTCATCAGAGTGAACAAGCAACCTACAGATGGGAGAAAAATTTTGGCAAT 7010
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1023 -LysGlnThrThrIleArgValAsnArgGlnProThrThrTrpGlnLysIlePheAla 1042
QY 7011 CTATCGATCTGCAAAAGGCTAATATCCAGATCTTACGAAGACTTAACAAATTTTCAAA 7070
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1042 rTyTrSerSerAspLysGlyLeuIleSerArgIleTyAsnGlnleuLysGlnIleTyLy 1062
QY 7071 GAAAAA-----AACACACCCGCTCAAAATATGAGCAAAAGATATGAGACACTTCGAAA 7126
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1062 sTyLysTrpAsnAsnProIleLysLysTrpAlaLysAspMetAsnTrpHisPheSerLys 1082
QY 7127 AGAAGCAATTATGACAGCCAAACAACATATGAAAAAACCTCATCATCTGCTGCTTAG 7186
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1082 sGlnAspIleTyAlaThrLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTGCATTCATCTCATCTGCTTGAATGGTGATCAC 7246
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1102 gGluMetGlnIleLysThrThrMetAlaTyTrpHisLeuThrProValArgMetAlaIleIl 1122

```

```

QY 7247 TAAAAAGTCAGAAACAAACAATGCTGGAGAGATGTGCAAAATAGAACACTTTTCCA 7306
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1122 eLysLysSerGlyAsnAsnAsnArgCysTrpArgGlyCysGlyGlnIleGlyThrLeuEHL 1142
QY 7307 CTGTGGTGGGAATGTAATTAATTAATGCTACACATTTGTGGAGACAGTGTGGAGATTCCTTAA 7366
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGAATCTGAAACCAAGAAATATCATCTTGGACCCAGCAATCCCAATTCAGTATATACCCAAA 7426
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1162 gAspLeuGlnleuGlnIleProPheAspProAlaIleProleuLeuGlyIleTyTrpAs 1182
QY 7427 GCAATATAATTCATCTATTAATAAGACACATGACACATATGTTATTCACAGACTGAT 7486
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1182 ngLutTyLysSerCysCysTyTrpLysAspTrpCysThrArgMetPheIleAlaLeuPh 1202
QY 7487 CACAAATGCAAAAGCTTGGAAACCAACCAAAATGTCATCAGTGTAGACTGTGATAAGAA 7546
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLy 1222
QY 7547 AACATGGCAGATATACACCCAAATGAAATCTATGACCCCATTAAGG-GATGAGTTCATGTC 7605
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1222 sMetTrpHisIleTyThrIleThrMetGlnTyTrpAlaAlaIleLysAsnAspGlnPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGATGAAGCTGGAACCATCATCTTCAGCAAACTAACACAAAGACA 7665
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1242 rPheValGlyThrTrpPheTyrLysleuGlnThrIleIleLeuSerLysLeuSerGlnGlnG 1262
QY 7666 GAAACCAAAACCCACATGTTCTCAGTTGTGAAGTGGAGAT 7705
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

```

## RESULT 9

```

ID Q15606 PRELIMINARY: PRT: 1275 AA.
AC Q15606:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel..01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dombroski B.A.;
RL submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-1275 FROM N.A.
RA Woods-Samuels P.;
RL submitted (APR-1989) to the EMBL/GenBank/DBJ databases.
DR EMBL: M80343; AAB59368.1; -.
DR EMBL: M22333; AAB88037.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RYTse.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR Pfam: PF00078; tvc; 1.
KW Hypothetical protein; RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 148971 MW; 6DA1229742848323 CRC64;

```

## Alignment Scores:

```

Pred. No.: 0 Length: 1275
Score: 5176.00 Matches: 1054
Percent Similarity: 78.83% Conservative: 78
Best Local Similarity: 73.40% Mismatches: 138
Query Match: 9.89% Indels: 173
DB: 4 Gaps: 7

```

US-10-083-853b-2 (1-29921) x Q15606 (1-1275)

```

QY 3428 ATGGCAGGATCAAAATTCACACATTAATTAATTAACCTTAATGTAATGGGCTAAATTC 3487
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 1 MetThrGlySerAsnSerHisIleThrIleLeuThrIleAsnIleAsnGlyLeuAsnSer 20  
QY 3488 CCAATTAAAGACACAGACTGGCAATTTGGATAAAGATCCAGACCATCAGTGTGCT 3547  
Db 21 AlaIleLysArgHisIleArgLeuAlaSerTrpIleLysSerGlnAspProSerValLysCys 40  
QY 3548 ATTACAGAGGCCCATCTGCATGAAAAAGACACACATAGGCTCAAAATTAAGGATGGAGC 3607  
Db 41 IleGlnGluThrHisLeuThrCysArgAspThrHisIleArgLeuLysIleLysGlyTrpArg 60  
QY 3608 AAGATTACCAAGTAATGGAAAAACAAAAAAGAGAGGAGGCTTCAATCCAGTCTCT 3667  
Db 61 LysIleTrpGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79  
QY 3668 GATAAACAAGACTTTAAACCAACAAGATCAAAAGACAAAGAGCCCATCATATAG 3727  
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlnHisIleMet 99  
QY 3728 GTAAAGGCATCATGACAGAAAGAGCTACATCTTAATATACATGACCCCATATCA 3787  
Db 100 ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTrpAlaProAsnThr 119  
QY 3788 GGAGCACCACAGATTCATAAGCAAGTTCTTAGAGACTACAAAGAGACTTTGACTCCAC 3847  
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgLysLeuAspSerHis 139  
QY 3848 ACAATATAGTGGAGCTTAATATTAATATAGACACTTTACACCCCACTGCCAATATTA 3907  
Db 140 ThrLeuIleMetLys-----AspPheAsnThrProLeuSerThrIle 153  
QY 3908 GGCAGATCATGACAGAAATTAATACAGGATATCCAGAGTGAAGTGAAGCTGTGCAC 3967  
Db 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173  
QY 3968 CAAGCGACCTATAGATATCTACAGAACTCCGCCCAATCAACAGATATACATC 4027  
Db 174 GlnAlaAspLeuIleAspIleLysArgThrLeuHisProLysSerThrGlyTrpPhe 193  
QY 4028 TTCTCGACATCACATTTACACCTATTAAATTAATGACCATGTAAATTTAAAGTAACACTC 4087  
Db 194 PheSerAlaProHisIleThrLysSerLysIleAspHisIleValLysSerLysAlaIle 213  
QY 4088 CTCAGCAATGCAAAAGACAGAAATCTTACAAAGATCTCAGACTACAGTACATGCAATC 4147  
Db 214 LeuSerLysCysLysArgThrGluIleThrAspThrLeuSerAspHisSerAlaIle 223  
QY 4148 TATTAGAACTCAGAAATTAAGAAATCTCACTCAAAATACACAACTACATGGAAGTGAAC 4207  
Db 234 LysLeuGluLeuArgLysIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn 253  
QY 4208 AACCTGCTCTGATATGATACATCTACGCGTAAATTAACAAATGAAGCAAAATTAAGATGTC 4267  
Db 254 AsnLeuLeuLeuAsnAspTrpTrpValHisAsnGlnMetLysAlaIleLysMetPhe 273  
QY 4268 TTGGAACCAATGAGAAACAAGACACATATGATACCAATCTCTGGGCAATATTAAAGCA 4327  
Db 274 PheGluThrAsnGlnLysAspThrTrpTrpGlnAsnLeuTrpAspAlaPheLysAla 293  
QY 4328 GTGTGTAGAGGAAATTTATAGCATAAGTATGCTTACAGAGAAAGAGGAGAAATATATAA 4387  
Db 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTrpLysArgLysGlnGluLeuArgSerLys 313  
QY 4388 ATAGACACCTTAAATCACAATTAAGAAAGACTAGAGAAAGAGCAAAACAATTAATAA 4447  
Db 314 IleAspThrLeuThrSerGlnLeuLysGlnLysGlnGlnIleThrHisSerLys 333  
QY 4448 GCTACAGAGACAGAAATTAATACTAATAGATACAGAGCAAGTGAAGATTAAGACACA 4507  
Db 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGlnIleGluThrGln 353  
QY 4508 AAAGCCCTCAATTAATCAATGAATCCAGAGCTGTGTTTGAAGAAATCAGCAAAAT 4567  
Db 354 LysThrLeuGlnLys-IleAsnGlnSerLysSerTrpPheGlnArgIleAsnLysIle 373

QY 4568 ----AGACCATAGACAGACTAATTAAGAAAGAGAGAGATCAAGATCAAT 4623  
Db 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgLysLysAsnGlnIleAspThrIle 393  
QY 4624 AAAAATGATTAAGGGGATATCCACCCAGTCCACAGAAATCAAACTTTCTGAGAG 4683  
Db 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgIle 413  
QY 4684 ATATTATTAACACCTCTATGCAATTAAGTAACTAGAAATCTAGAGAAATGATTAATCTCT 4743  
Db 413 uTrpTrpLysHisLeuTrpAlaAsnLysLeuGlnLysLeuGlnGluMetAspPheIle 433  
QY 4744 GCACACATATGTAGCTGTATGACCTTGGGGACACAAACAAAGGGGATGATGCAGAA 4803  
Db 433 uAsp----- 434  
QY 4804 ATAAAGACAAAGACAAAGATATGTTTGAAGTAGGGGTCAAGGGGCAACTGTGCTCT 4863  
Db 434 ----- 434  
QY 4864 AATGACAAAGGCCCTGAGCTTTACACACACCTCTGATTTATTAAGCAAAAGATAGAC 4923  
Db 434 ----- 434  
QY 4924 GAGAGGGTAGTTGGAAGAAGAGTACGCTGTAGTCCAGAGTACGCTGCAAGACTGC 4983  
Db 434 ----- 434  
QY 4984 ATTCTCAAAACATAGGCTCTAGATGTCACAGTAGATTAACCTCAAGAGCAGTGCAGG 5043  
Db 434 ----- 434  
QY 5044 GAGTATGGCCCTCAGCAAACTCTTAGAGGACAGACAGAAATTAAGTTGCCACATCT 5103  
Db 434 ----- 434  
QY 5104 GTATTACGATTAACAGTTTGCCTTGTGATCAAGTAGCCTCCAGTGAATGCTGAATGG 5163  
Db 434 ----- 434  
QY 5164 TCATGATCCCTTGGCTTTTGGCTCCCAAAACATACACCTCTCAAGACTAAACCA 5223  
Db 435 -----ThyTrpThrLeuProArgLeuAsnGln 443  
QY 5224 GGAAGAAGTCAAATCCCTGAATATACCAAGTACCAAGTCTTAAATTAAGACGTAATGA 5283  
Db 443 nGluGlnValGlnSerLeuAsnArgProIleThrGlySerGlnIleValAlaIleIleAs 463  
QY 5284 TAGCTTACCAACCAAAAGATCCAGACAGAGGATTCACAGCCAAATTCATACAGAG 5343  
Db 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTrpGln 483  
QY 5344 GTACAAAGAGAGCTGTACTATTCTTGTGAACATATTCACAAATAAGAA---AATG 5400  
Db 483 gTryMetGlnGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGlnLysGln 503  
QY 5401 GAATCTCTCTAATCATTTTACAGAGCCAGCATCATCTGATACCAAAACCTAGCAGTG 5460  
Db 503 lYleLeuProAsnSerPheTrpGluAlaSerIleIleLeuIleProLysProGlyArg 523  
QY 5461 ACACAACAAAGAGAAATTTTCAGGCCCATATCTCTGATGAACATTTGATGTGAATAATC 5520  
Db 523 sPThrThrLysLysGlnAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIle 543  
QY 5521 TCAATTAATTAATGCGCAACCAATTCACAGACATCAAAAGCTTATCTACCAATG 5580  
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG 563  
QY 5581 AAGTTGGCGTCATCCCTGGATGCAAGGCTGTCAAAATATGCAATCAATCAATATGTAG 5640  
Db 563 InvalGlyPheIleProGlnMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 583

QY 5641 GCCATCACAATAAGACAAATGACAAAAACCATGATTAATTCATAGATGCAGAA 5700  
|||||  
Db 583 IGLNIIITLSEANATGALATASPLYSASNHISETLLEISERILEASPLAIGLU 603  
QY 5701 AGGCCCTTTCAAANATCAACAGCCCTTTCATTAATAATTCAGTAACATAGATG 5760  
|||||  
Db 603 YSALPHEASPLYSILEGInGInProPhemeLeuysThrLeuSnLysLeuGlyLea 623  
QY 5761 ATGGAATGTATCTCAAAAATTAAGAGCATTTATAC-AAACCCACAGCCATATATAC 5819  
|||||  
Db 623 SPGLYThrThyPheYsIleIleHgaIleIleTyraSplysProThrIlaSnIleIle 643  
QY 5820 TGAATGGCCAAAAGTGAAGCATTCCTTTGAGAAGTGCACACAGACAGATGCCCTC 5879  
|||||  
Db 643 EUSANGLYGLInLysLeuGluAlaPheProLeuLysThrGlyThraGngInLysProL 663  
QY 5880 TCTCACACTCCTATTCAAGATACATTTGGAAGTTCTGGCCAGGCAATCAGCAATGAC 5939  
|||||  
Db 663 EUSerProLeuLeuPheAsnIleValLeuGluValLeuAlaIleArgAlaIleArgInGlu 683  
QY 5940 AAGAATAAAGGTATTCATAATAGAAAGAGAGAGATCATATTGCTCMTTGCAGATG 5999  
|||||  
Db 683 YSGInLysGlyIleArgLeuGlyLysGluGluValLysLeuSerLeuPheAlaSPH 703  
QY 6000 ACATGTTTGTATATTAGAAAACCCCATGCTCTCAGGCCAAAACCTCTTAAGCTGATA 6059  
|||||  
Db 703 SPMeIleValIleYrLeuGluAsnProIleValSerIleGlnsInLeuLysLeuIleS 723  
QY 6060 GCAACTTCGCAAAAGTCTCAGACACACAAAATCAATGSCAAAATACACAGCATTTCTAT 6119  
|||||  
Db 723 ERASpHeSerLysValSerGlyTyLysIleAsnValGlnLysSerGlnAlaPheLeuT 743  
QY 6120 ACGCCAAATATAGACAAAACAGAGAGCAATCATGATGTCATCACAATGTGTA 6179  
|||||  
Db 743 YrThraSnsnArgInnHrgLusSerGlnIleMeGlyLysLeuProPheValIleAlaS 763  
QY 6180 CAAAGAGATTAATAATACCTAGAGATCAACTTACAGAGGACAGTGAAGTCTTCAAG 6239  
:::|||||  
Db 763 ERlysaIrgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783  
QY 6240 AAGACTACAACCACTGATCAAGGAATATAGAGAGACACAAACAAATGGAACATTC 6299  
|||||  
Db 783 IASnTyLysProLeuLeuLysGlnIleLysGlnLysPThrAsnLysTyLysAsnIleP 803  
QY 6300 CATGCGACACAGATAGATCAATCAT-----GAATAATGCTATCGCCCAAGTAAT 6352  
|||||  
Db 803 ROCYSSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823  
QY 6353 ATAGATTCAGTCTACCCCATCAACCTACATTCATCTTCTTCACAGATTTGGAAGAA 6412  
|||||  
Db 823 YrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT 843  
QY 6413 CAACCTTAATTCATATGGAACCAAAAGAGCCACAGAGCCAGCAATCTTTAAGC 6472  
|||||  
Db 843 HrrThLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862  
QY 6473 AAAAAGAACAAAGCTGAGATTCATGATCAGCTGAGCTTAATAACTATACATTAAGCTACA 6532  
:::|||||  
Db 863 GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyTyLysAlaThr 882  
QY 6533 GTAACCAAAAAGTGCATGTACTGTGACCAAAACAGATATATAGACCAATGGAACAGACA 6592  
|||||  
Db 883 ValThrLysThrAlaIleTrpTyTrpYrGlnAsnArgAspIleAspInnTrpAsnArgThr 902  
QY 6593 GAGACCTCGAAGATATAC-ACATGCAATCTACATCCATCTGATCTTTGACAAACCTGACAA 6651  
|||||  
Db 903 GluProSerGluIleMetProHisIleTyraSnTyLysIlePheAspLysProGluLys 922  
QY 6653 AACACAGCAATGGAAGAAAGATCCCATTTAATTAATGTTGTAAGAAAGCGCATGCC 6711  
|||||  
Db 923 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpLysInnTrpLeuAla 942  
QY 6712 ATATCGAAGAAAGCTGAAGCTGATCCCTTCTTACACCTTATACAAAAGTTAATCAAGA 6771

Db 943 ILeCysArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAsnSerArg 962  
QY 6772 TGAATTAAGACCTTAATATTAAGACATTAACCATTAACCA-CAAGAAAACCTGAGC 6830  
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGlnGlnsInLeuGly 982  
QY 6831 AATACCATTCAGCATATGACATGAGCAGCAAGACTTCATACATTAACACCAAAAGCAAG 6890  
|||||  
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerIleTyThrProLysAlaMet 1002  
QY 6891 GCAACAAAAGCCAAAATAGACAGATGGATCTGATTAACATTAAGAGCTTTCACAGACA 6950  
|||||  
Db 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022  
QY 6951 AAAAAAAGTGTATCAGAGTGAACAGACAGCAACTACAGATGGGAGAAAAATTTTGCAAT 7010  
|||||  
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrTrpGluLysIlePheAlaTh 1042  
QY 7011 CTATCATCTGACAAAGCTTAATATCCAGAGATCTACGAAAGCTTAACAAATTTACAA 7070  
|||||  
Db 1042 rTySerSerAspLysGlyLeuIleSerArgIleTyraSngLysLeuLysGlnIleTyTy 1062  
QY 7071 GAAAAA----AACACCCCGTCAAAAATATGGCCAAAGATATAGACAGACACTTTCAAA 7126  
|||||  
Db 1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAlaAsnArgHisPheSerLys 1082  
QY 7127 AGAAGCATTTTATGACGCCACAAACATATGAAAAAAGCTCATCATTTGCTGCTAG 7186  
|||||  
Db 1082 sGluAspIleTyraIleAlaLysLysHisMetLysCysSerSerSerLeuAlaIle 1102  
QY 7187 AGAATGCAAAACAAACACAGCTGACATACCTTCATCGCTGATTAAGATGGTGAATCAC 7246  
|||||  
Db 1102 gGluMetGlnIleLysThrThrMetLargYrHisLeuThrProValArgMetAlaIleI 1122  
QY 7247 TAAAAAGTCAGAAACAAACAAATGCTGAGAGAGATGGAAGATAGGAACATTTCACA 7306  
|||||  
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgLysGlyGluIleGlyThrLeuLeuH 1142  
QY 7307 CTGTGCTGGGAATGTAATTAATTAATTCATCATTTGTGGAAGACAGTGTGAGATTCCTTAA 7366  
|||||  
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuTr 1162  
QY 7367 GGAATGCAAGCCGAATATATCATTTGACCCAGCAATCCCATTCGATGATTAATACCAAA 7426  
|||||  
Db 1162 gaspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyTrpAs 1182  
QY 7427 GGAATTAATCATTTCTATTAATAAGACACATGACACATATGTTTAATGACACACTGAT 7486  
|||||  
Db 1182 nGluTyLysSerCysGlyTyLysAspThrCysThrArgMetPheIleAlaIleLeuPh 1202  
QY 7487 CACAATAGCAAGACTGTGACCAACCAATTCATCATGATGATGACGTGATAAGAA 7546  
|||||  
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLys 1222  
QY 7547 AACATGCAATATACACATGATTAATGATGAGCCATTAAG-GATGAGTTCATGTC 7605  
|||||  
Db 1222 sMetTrpHisIleTyThrMetCylutyTyraIleAlaIleLysAsnAspGluPheIleSe 1242  
QY 7606 CTTTGCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665  
|||||  
Db 7666 GAAACCAACACACATGTTCTCATCTGTAAGTGGAGAT 7705  
|||||  
Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyLysAsn 1275  
RESULT 10  
000363 PRELIMINARY; PRT: 1275 AA.  
AC 000363;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, last sequence update)







QY 5164 TCATGATCCCTTGGCTTTTGGCTCCCAAAACATACACCTCTCAAGATAACCA 5223  
Db 435 -----ThrTyrThrLeuProArgLeuasnG1 443  
QY 5224 GGAAGAAGTCAAATCCCTGATATACCAAGCTCTAAATGAGCACTAATGA 5283  
Db 443 ngIuGluValGluSerLeuasnArgProIleThrGlySerGluIleValAlaIleIleAs 463  
QY 5284 TACCACACCAACAAAAAAGTCCAGACAGCGAGTATTCACGCCAAATTCACAGAG 5343  
Db 463 nSerLeuProThrLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnAr 483  
QY 5344 GTCAAGAAGAACTGGTACTATTCTTGAAGCATTCACAAAAATAAGAA---AATGG 5400  
Db 483 gTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu- 503  
QY 5401 GAATCTCTCCATACATTTTACGAGCCAGCATCATCTGATACCAAAACCTAGCAGTG 5460  
Db 503 LysLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArgA 523  
QY 5461 ACACACCAAAAGAGAAATTTGACGCCCATATCCCTGATGAGCACTTGAATGTAATCC 5520  
Db 523 sPthrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543  
QY 5521 TCATAAAATACATGCGAAACCAATCCAGCAGACATCAAAAGCTTATCTACATGATC 5580  
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnIleIleLysLysLeuIleHisHisAspG 563  
QY 5581 AACTGGCGTCATCCCTGGGATGCAAGGCTGTCAAAATATGCAAAATCAATTAATGTAG 5640  
Db 563 InValIleGlyPheIleProGlyMetGlnGlyTyrPheAsnIleArgLysSerIleAsnValI 583  
QY 5641 GCATACACATTAACACCAATGACAAACCAACATGATATCTCAATGATGACAGAAA 5700  
Db 583 IeGlnIleHisAsnAlaGalaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 603  
QY 5701 AGGCTTTTGTCAAAATTCACACGCTTCATGCTAAAAATCTCAGTAACTAGTATCG 5760  
Db 603 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623  
QY 5761 ATGGAATGTATCTCAAAATATTAAGAAGCTATTTATAC-AAACCCACACCAATATCATAC 5819  
Db 623 sPylMetLysTyrPheLysIleIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL 643  
QY 5820 TGAATGGCAAAACGGAAGCATTCCTTGTGAGACTGGCACAAGCAAGATGCCCTC 5879  
Db 643 euAsnGlnLysLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 663  
QY 5880 TCTCACCACTCTATTCAGATATCTATTGGAAGTCTGGCCAGGCAATCAGCAATAGA 5939  
Db 663 euSerProLeuLeuPheAsnIleValLeuGlnValLeuAlaArgAlaIleArgGlnLul 683  
QY 5940 AAGAATAAAGGATTCCTCAATAGAAAGAGAGAGAGCATATATGCTCTGTTTGCAGATG 5999  
Db 683 ysgIuIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 703  
QY 6000 ACATGTTTGATATTAGAAACCCGATGCTCAGGCCCAAAACCTTAAACCTGATAA 6059  
Db 703 sPheIleValIleValLysLeuAsnProIleValIleSerAlaGlnAsnLeuLeuLysLeuL 723  
QY 6060 GCAACTTCAGCAAGCTCTCAGACACAAATCAATGTGCAAAATGCAACAGATTTCTAT 6119  
Db 723 ersnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeuT 743  
QY 6120 ACCGCAATATAGACAAACGAGAGCCAAATCATGAGTGAAGTCTCATTCACAATGCTGA 6179  
Db 743 yThrTrsAsnAsnArgGlnThrGluSerGlnIleMetGlyLysLeuProPheThrIleAl 763  
QY 6180 CAAGAAGTAATTAATACCTGAATCAACTTACAGGAGCAGGTAGAGACTCTTCAGG 6239  
Db 763 erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783

QY 6240 AGAATACAAACCACTGATCAAGAAATAGAGAGGACACAAACAAATGAAAAATTC 6299  
Db 783 IuAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTyrLysAsnIleP 803  
QY 6300 CATGCTCAGATAGTAGAATCAT-----GAAAATGCCATATCTGCCAAAGTAATT 6352  
Db 803 rocYsSerTyrValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823  
QY 6353 ATGATTCATGCTGATCCCATCAAGTACCATTTGACTTCTTCACAGAATTTGAAAAA 6412  
Db 823 yThrPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT 843  
QY 6413 CAACTTAAATTCATATGAGAACCAAAAAAAGAGCCACAGACCAAGACAACTTATAGC 6472  
Db 843 hrThrLeuLysPheIleThrPasnGln-LysArgAlaArgMetAlaLysSerIleLeuSer 862  
QY 6473 AAAAAGAACAACTGGAGGTATATGCTATCTGACTTAAACTATATCTAATAAGCTACA 6532  
Db 863 GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyrLysAlaThr 882  
QY 6533 GTAAACCAAACTGATGCTGCTAGTACCAAAACAAATATATACCAATGACACAGACA 6592  
Db 883 ValThrLysThrAlaTyrTyrTyrGlnAsnArgspIleAspGlnTyrPheArgThr 902  
QY 6593 GAGACCTCAGAAATTTAC-ACCTGCAATCTACATCATCTGATCTTGTGCAAACTGACAAA 6651  
Db 903 GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys 922  
QY 6652 AACACGCAATGGAAGAAAGATTCCTTATTAATATGATGTTGGAAAAACCTGGCTAGCC 6711  
Db 923 AsnLysGlnTyrPheLysAspSerLeuPheAsnLysTyrPysTyrGluAsnTyrLeuAla 942  
QY 6712 ATATGACGAAGCTGAAACGATGCCCTTCCTTACACCTTATCAAAAGTATCTCAGAGA 6771  
Db 943 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg 962  
QY 6772 TGAATTAAGACTTAAATATATAGACATAAACCATAAACCA-GAAGAAAACTTAGGC 6830  
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly 982  
QY 6831 AATACCATTCAGATATGACATGGGCAAAAGCTTCATGACTTAAGAACCAAAAGCAATG 6890  
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002  
QY 6891 GCAACAAACCCAAATAGACAGTGGGATGCTATTAATATAGACTCTGCACAGCA 6950  
Db 1003 AlaThrLysAspLysIleAspLysTyrPaspLeuIleLysLeuLysSerPheCysThrAla 1022  
QY 6951 AAAAAAACTGTCTATCAGATGAGTGAACAAGCACTACAGATGGGAGAAAAATTTTGCAT 7010  
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrThrTyrGluLysIlePheAlaThr 1042  
QY 7011 CTATCGATCTGACAAAGCTAATATATCCAGAGATCTACGAAGAACTTAACAAATTTACA 7070  
Db 1042 rTyrSerSerAspLysGlyLeuLysLeuSerArgIleTyrAsnGluLeuLysGlnIleTyrL 1062  
QY 7071 GAAAAA-----AACACCCCGTCAAAATATAGGGAAGATAGGAGACAGACTTCGAAA 7126  
Db 1062 sLysLysThrAsnAsnProIleLysLysTyrValLysAspMetAsnArgHisPheSerL 1082  
QY 7127 AGAAGACATTTATGACACCAACAACATATAGAAAAAACTCATCATCTATGCTGTAG 7186  
Db 1082 sGluAspIleTyrAlaIleLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr 1102  
QY 7187 AGAAATGCCAAACAAAACCAAGTGCATATCCATCTGATCTGATTAAGATGGTATCAC 7246  
Db 1102 gGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI 1122  
QY 7247 TAAAGATCAGAAACCAAAATGCTGAGAGATGAGTGGAGAAATGGAACATTTTCCA 7306  
Db 1122 eLysLysSerGlyAsnAsnArgCysTyrPargGlyCysGlyGluIleGlyThrLeuLeuH 1142  
QY 7307 CTGTGGTGGGAATGTAAATTAAGTCAACCATTTGTGGAAGACAGTGTGAGATTCCTTAA 7366

```

Db      1142 scystriprrpascyslysluvaljglnproleutrrpsserValtrprrgheleuar 1162
OY      7367 GGAATTCAGAACCAAGAAATATTCATTGACCAGCAATCCATTACTAGATATATACCAAA 7426
Db      1162 gnspleugluleuglulleprophesproalalleproleuenglylelyrrproly 1182
OY      7427 GGAATTAATATCTTTATTTATTAAGACACATGCACACATATGTTTATTGACACTGAT 7486
Db      1182 sasprtyrlyssercyscyslyrrysasprhrcysttrrargmetPheillealaleuuh 1202
OY      7487 CCAATAGCAAGACACTGGACCAACCAATGTGCATGATGATGATGATGATGATGATGATGAT 7546
Db      1202 ethrillealalystrtrpshnclnprolyscsprottrmetilleasprtrilleysly 1222
OY      7547 AACATGCACATATACCCATGAAATCTATGCAGCCATAAAAG-CATGAGTTCATGTC 7605
Db      1222 shetrtrpshlletytrthmetclutyrtrrhaalalleysasnsapglupheleese 1242
OY      7606 CTTTGACAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7665
Db      1242 rphevalgllytrtrmetclyleuglunthrillelleuSerlyleuSerGlnlucgl 1262
OY      7666 GAAACCAACACACACATGTTCTCACTGTGTAAGTGGAGT 7705
Db      1262 nlystrlyshlsargllepheserleulleglylyasn 1275

```

## RESULT 11

```

012881 PRELIMINARY; PRT: 1275 AA.

```

```

AC 012881;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Retrotransposable Ll element LRE2 from chromosome 1q.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN NCB1
RP SEQUENCE FROM N.A.
RC TRANSPOSON-LRE2;
RA MEDLINE=95004577; PubMed=7920631;
RA Holmes S.E., Dombroski B.A., Krebs C.M., Boehm C.D., Kazanian H.H.;
RT "A new retrotransposable human Ll element from the LRE2 locus on
RT chromosome 1q produces a chimaeric insertion.";
RL Nat. Genet. 7:143-148(1994).
DR EMBL: U09116; AAB60345.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF03372; Exo_endo_phos: 1.
DR Pfam: PF00078; rvc: 1.
KW RNA-directed DNA polymerase.
FT VARIANT 1182 1182 E -> K.
SQ SEQUENCE 1275 AA; 149064 MW; CEF4429261FE223 CRC64;

```

## Alignment Scores:

```

Pred. No.: 0 Length: 1275
Score: 5169.00 Matches: 1054
Percent Similarity: 78.908 Conservative: 79
Best Local Similarity: 73.408 Mismatches: 137
Query Match: 9.888 Indels: 173
Gaps: 7

```

```

US-10-083-853b-2 (1-29921) x 012881 (1-1275)

```

```

OY      3428 ATGGCAGGATCAATTCACACATTAATTAACCTTAATGTAATGGCTAAATTC 3487
Db      1 MettrnglySerAsnSerHisIleThrIleuThrIleuAsnIleAsnIlyleuAsnSer 20
OY      3488 CCAATTAAGACACAGACATGCAATGATTAAGAGCAAGACACACCATGAGTGTCT 3547
Db      21 AlalleysArgHlsArgLeuAlaSerTrpIleIlysSerGlnAspSerValCysCys 40

```

```

OY      3548 ATTACAGAGCCCATCTCACATGAAGACACACATAGCTCAAAATTAAGGATGAGG 3607
Db      41 lIeGlnGlutHrHisIleuMetCysArgAspThrHisArgLeuIlysIleuSgIlyrrArg 60
OY      3608 AAGATTACCACTAAATGGAACCAAAAAAAGCAAGGGGTTCATTCCTAGCTCT 3667
Db      61 LysIleTyrGlnIleAsnGlyLysGln--LysLysAlaGlyAlaAlaIleuValSer 79
OY      3668 GATTAACAGACTTTTAACCAACCAAGATCAAAAGAGCAAGAGCCATTAACATATG 3727
Db      80 AsprlyThrAspPheIysProThrLysIleIysArgAspLysGlnIlystIleMet 99
OY      3728 GTAAGGCATCAATGGAACAGAGAGCTACTATCTTAATATATACATGACCCCAATTA 3787
Db      100 VallyGlySerIleGlnGlnGlnIleuThrIleLeuAsnIleTyrAlaProAsnThr 119
OY      3788 GAGACCCCAAGATTCATTAAGACAGAGTCTTAGAGACTTAAAGACACTTGACTCCAC 3847
Db      120 GlyAlaProArgPheIleLysGlnValIleuSerAspLeuGlnArgAspLeuAspSerHis 139
OY      3848 ACAATATAGTGGAGCTTAATTAATATATAGACACTTAAACACCCCACTGCCAATATTA 3907
Db      140 ThrIleIleMetGly-----AspPheAsnThrProIleuSerIleIleu 153
OY      3908 GCGAGATCAATGAGACAGAAATTAACAGATATCCAGAGTGAAGTGAAGCTGTGAC 3967
Db      154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnIleuAsnSerAlaIleuHis 173
OY      3968 CAAGCGGACCTAATAGATATCTACAGAACTCCCAACCCCAATCAACAGATATACAC 4027
Db      174 GlnAlaAspLeuIleAspIleTyrArgThrIleuHisProLysSerThrIlyThrPhe 193.
OY      4028 TTCTCAGATCACTATACCTATTTTAATTAATGACATGATATTAATTAAGTAACATC 4087
Db      194 PheSerAlaProHisHisThrIlySerIlyIleAspHisIleValGlySerIlyAlaIleu 213
OY      4088 CTCACGAAATGCAAAAGAACAGAAATCTTAACAAACAGTCTCTCAGACTACAGTGCAT 4147
Db      214 LeuSerIlyCysLysArgThrGlnIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
OY      4148 TATTTAGACATGAGATTAAGAACTCAACTCAACACAACTAATGGAAGACGAC 4207
Db      234 LysLeuGlnIleuArgIleLysAsnIleuThrGlnSerArgSerThrThrTrpIlyLeuAsn 253
OY      4208 AACCTGCTCTCAATGACTACTGCGTAATTAACAATGAAGCAAAATTAAGATGTC 4267
Db      254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGlnMetLysAlaGlnIleLysMetPhe 273
OY      4268 TTGGAACCAATGAGAACAAAGACACATGACCAAGATCTCGGGCATATTTAAAGCA 4327
Db      274 PheGlnThrAsnGlnAsnLysAspThrTrpTyrGlnAsnIleuThrAspAlaPheLysAla 293
OY      4328 GTCTGAGAGGGAATTTATACACATGAGTGGCTCAAGAGAAAGACGAGAATTAATTA 4387
Db      294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnIleuArgSerLys 313
OY      4388 ATAGACACCTTAACATCAACATTAAGAAAGACTAGAGAAAGAGCAAAACAATTCAAA 4447
Db      314 IleAspThrLeuThrSerIleuLysGlnIleuGlnIlyGlnGlnIleuThrHisSerLys 333
OY      4448 GCTAGCAGAGAACAGAAATTAACATAGATCAGACGAAAGCTAGAGAGATAGACACAAA 4507
Db      334 AlaSerArgArgGlnGlnIleThrLysIleArgAlaGlnIleuLysGlnIleGlnIleuThrGln 353
OY      4508 AAGCCCTTCAATTAATCAATCAATCCAGAGCGCTGTTTGAAGAAAGATCAGCAAAAT 4567
Db      354 LysThrLeuGlnLys-IleAsnGlnLysSerTrpPhePheGlnValArgIleAsnLysI 373
OY      4568 ----AGACACATAGACAGACTAATTAAGAGAAAGAGAGAGAGATCAAGAGATGACAT 4623
Db      373 eAspArgProLeuAlaIleArgLeuIleLysLysArgGlnIlyLysAsnGlnIleAspThrI 393

```

QY 4624 AAAAAATGATTAAGGGATATTCACACCGATCCACAGAAATACAAATATATACAGAGA 4683  
|||||  
Db 393 eutsasnasplysGlyAspIlethrhrasprothrInglIleGlnThrIleatrgl 413  
QY 4684 ATTTTAAACACCTCTATGCAATTAACAGAAATCTGAAGAAATGATTAATTCCT 4743  
|||||  
Db 413 uTyTyTyShsIleuTyTAlasInLysLeuGlnuLeuGlnIleuMetAspThrPhele 433  
QY 4744 GGACACATATATGCTGTATGACCTTGCGGGACAGAACAAAAGGGGATGACGAA 4803  
|||||  
Db 433 unsp----- 434  
QY 4804 ATTAAGAGACAAAGACAAAGATATGTTGGAAGTAGGGGTGAGGGGCAACTTGCTCT 4863  
434 ----- 434  
QY 4864 AATGCAAGAGGCGCTGAGCTTACACACCCTCTGTATTATTATTAAGCAAAAGATAGC 4923  
434 ----- 434  
QY 4924 GAGAGGTGAGTTGGAAGAAGAGGTGACAGTGTAGGTCGAGAGTGGCCGCAAGACTGC 4983  
434 ----- 434  
QY 4984 ATTCTCAACAATAGCTCTAGATGTCCAGTAGATTAACCTCAAGAGCCAGTGCAGG 5043  
434 ----- 434  
QY 5044 GAGTGTGGCCCTGACCAAACTTCTAGGCGACAGGACAGAAATAGTTTCCCACTTCT 5103  
434 ----- 434  
QY 5104 GTATTCACGATAAACAATTGTCTGTGTGATCAAGTAGCTCCAGTGGAAATGCTGAGTTGG 5163  
434 ----- 434  
QY 5164 TCATGATCCCTTTGGCTTTTGGCTCCCAAAACATACACCTCTCAAGACTTAACCA 5223  
435 ----- 435  
QY 5224 GGAAGAGTCAATCCCTGATATACAGTAACAGATTTCAAAATGAGACAGTAATGA 5283  
|||||  
Db 443 ngIuGluValGluSerLeuasnngIyProIlethrIySerGlnIleValAlaIleIleas 463  
QY 5284 TACCTACGCAACCAAAAAAGTCAGAGCAGACGAGTTCACAGCCAAATATACAGAG 5343  
463 nserLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrglnat 483  
QY 5344 GTACAAAGAGAGCTGTACTATCTTCTGAAACTATTCCAAAAAATAGAA--AATGC 5400  
|||||  
Db 483 gTyTyLysGlnGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGlnLysGlu 503  
QY 5401 GAATCTCTCCCTACTATTTTACGAGCCGACATCATCCGATACCAAAACCTAGCGTG 5460  
|||||  
Db 503 LylleuProAsnSerPheTyrgluAlaSerIleIleLeuIleProLysProGlyArg 523  
QY 5461 ACACACAAAGAGAAATTTAGGCCCCATATCCCTGATGAACATGATGATGATAATTC 5520  
523 spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543  
QY 5521 TCAATAAAATCTGCAAAACCAATCCAGACGACATCAAAAAGCTTATCTACCATGATC 5580  
543 euaInLysIleLeuAlaAsnArgIleGlnGlnIleLysIleLysIleIleHisIleAsp 563  
QY 5581 AATTTGGCGTACCTCCGAGTGCAGAGCTGTGTTCAAAATATGCAATCAATAATAG 5640  
563 InValGlyPheIleProGlyMetGlnGlyTyrrPheAsnIleArgLysSerIleAsnValI 583  
QY 5641 GCATACATAAACAGACCAATGACAAAACCAACATGATTAATCAATGATGACAGAAA 5700  
583 IeGlnHisIleAsnArgAlaLysAspLysAsnHisIleIleSerIleIleAspAlaLul 603  
QY 5701 AGGCTTTGTCAAAATTCACAGCCCTTCACTGCTAAAAATTTCTAGTAACCTAGTATCG 5760

Db 603 ysaIlePheAspLysIleGlnGlnProPheMetLeuLysThrIleAsnLysLeuGlyIleA 623  
QY 5761 ATGGAATGTATCTCAAAATTAATAAGACTTTTATAC-AAACCCACAGCCCAATATCTAC 5819  
623 spGlyThrTyrrPheLysIleIleArgAlaIleIleTyrrAspLysProThrAlaAsnIleIle 643  
QY 5820 TGAATGGGCAAAACCTGAGCATTCCTCTGGAACCTGGCACAGCAAGAGATGCCCTC 5879  
643 euaInGlnGlnLysLeuGlnAlaPheProLeuLysThrGlyThrArgGlnGlySerIle 663  
QY 5880 TCTCACACCTCTATTCAGATACTATTTGGAAGTCTGCGCAGGCAATCAGCAATAGA 5939  
663 euaSerProLeuLeuPheAsnIleValLeuGlnValleuAlaArgAlaIleArgGlnLul 683  
QY 5940 AACGAATTAAGGCTATTCAAATGAAAAGAGAGAGTCAATTTCTCTGTTGCAGATG 5999  
683 yscIuIleLysGlyIleGlnLeuGlyLysGlnGluValLysLeuSerLeuPheAlaAsp 703  
QY 6000 ACATGTTTGTATATTTTGAAGAACCCCATGCTGAGCCCAAAAACCTTAAGCTGATA 6059  
703 spMetIleValIlyLeuGlnAsnProIleValSerAlaGlnAsnLeuLysLeuIles 723  
QY 6060 GCAACTTCACAAAGTCTCAGACACAAAATCAATGTGCAAAAATCACAAGATTTCTAT 6119  
723 eraSnPheSerLysValSerGlyTyTyLysIleAsnValGlnLysSerGlnAlaPheLeu 743  
QY 6120 ACCGCATATATACAAACAGACAGACCAATTCATGAGTGAACCTCTATTCACATTCCTA 6179  
743 yrrhrAsnaAsnArgInThrGlnSerGlnIleMetGlyGlnLeuProPheThrIleAlas 763  
QY 6180 CAAGAGAAATTAATACCTAGATACACTTCAAGGAGCAGCTAGCACTTCAAG 6239  
763 eTyTyAsnArgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLys 783  
QY 6240 AGAAGTACAACACCTGATCAAGAGAAATTAAGAGAGACACAAACAAATGAAAAACATTC 6299  
783 lAsnTyTyLysProLeuLeuLysGlnIleLysGlnAspThrAsnLysTyrrLysAsnIleP 803  
QY 6300 CATGCTCAGATATGATGAATCAT-----GAAATGCCATTAAGTCCCAAGTAAT 6352  
803 rocYsSerTyrrValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIle 823  
QY 6353 ATGAGATTCAGTGCACCCCATCAAGCTACCATGACTTCTTCACAGAAATGAAAAA 6412  
823 yrrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGlnLeuGlnLys 843  
QY 6413 CAATTTAAATTCATATGGAACCAAAAAAGAGCCACAGAGCAAGCAAGCAATCTTAAGC 6472  
843 hrThrLeuLysPheIleTyrrAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 862  
QY 6473 AAAAAGAACAAAGCTGAGTATCATCTACCTGACTTAACCATATCTAATAGGCTACA 6532  
863 GlnLysAsnLysAlaGlyGlyIleThrLeuProTyrrPheLysLeuTyTyLysAlaThr 882  
QY 6533 GTAACCAAACTGCATGCTAGTACCAAAAACAGATATATAGCAATGGAAGAGACA 6592  
883 ValThrLysThrAlaTyrrTyrrTyrrGlnAsnArgLysPheGlnIlePheAsnArgThr 902  
QY 6593 GAGACCTCAGAAATTC-ACGCAATCTACATCCATGTGATCTTTGACAAAACCTGACAAA 6651  
903 GluProSerGlnIleMetProHisIleTyAsnTyrrLeuIlePheAspLysProGlnLys 922  
QY 6652 AACACAGAAATGAAAAAGATTCCTTTTAAATAAAGGTGTGAAAAACTGGCTAGCC 6711  
923 AsnLysGlnTyrrGlyLysAspSerLeuPheAsnLysTyrrTyrrGlnAsnTyrrLeuAla 942  
QY 6712 ATATGCAAGAGGAGTGAATGATCCCTCTTACCTATATACAAAGTAACTACAGA 6771  
943 IleCysArgLysLeuLysLeuAspLeuPheLeuThrProTyrrThrLysIleAsnSerArg 962  
QY 6772 TGAATTAAGACTTAAATATAGACATAAAAACCAATAAAAACCA-GAAGAAAACCTAGGC 6830



OY	4148	TATTTAGAAGCTCAGAATTTAAGAACTCACCTCAAAAATCACACAACTACTGGAAC	4207
Db	234	LysLeugInIueuArGIleLysAsnIuehrGlnSerArgSerThrTrpLysIueasn	253
OY	4208	AACCTGCCCTCGAATGACTACTGGTAAATAACAATAAGAACGAAAAATTAAGTGTC	4267
Db	254	AsnIueIueSerAsPtyrTrpIuehisasnIueLysalaGlIueLysWerthe	273
OY	4268	TTTTGAAACAATGAGAAACAAGACACAATGTACAGCAATCTCTGGGCATATTAAACA	4327
Db	274	PheglunThrsangluasnlLysaspHrThrIyrGlnsnIueutTPaspAlaPhelysLa	293
OY	4328	GTCGTAGAGGAAATTTATGACACTAGATCCCTPACAAGAGAAACGAGAAATATCTAA	4387
Db	294	ValCysargGIlyLysPheilealaleuasnlatyrlLysargLysGlnIargSerlys	313
OY	4388	ATAGACACCTTACATCATCAATTTAAAGACTAGAGAAAGAAAGGCAACCAATTCAAA	4447
Db	314	IlespnhrIuehnSerGlnIueLysGlnIueGlnLysGlnIueGlnThrhiserLys	333
OY	4448	GCTAGCAGACAGACAAGAAATACTAAGATCAGAGCAGACGAACTGAAGAGATGAGACACA	4507
Db	334	AlaserSerArgIngluIlleThrylsIleaArglaGlIueLysGlnIleGluthrcIn	353
OY	4508	AAAGCCCTTCAAATTAATCAATGAATCCAGACCTGGTTTTTGAAGAAATCAGCAAAAT	4567
Db	354	LysThrIeugInLys-IleasnGlnSerArgIlletrPhepheGluArgIleasnLysIl	373
OY	4568	---AGACCCTAGACACTAATTAAGAAAGAAAGAGAAAGATCAACAGATGCAT	4623
Db	373	eAsprIrProIueuIahrGleuIIeLysLysLysArgelIuLysasnGlnlleasPrhlI	393
OY	4624	AAAAAATGATAAAGGGGATATCACCCAGCATCCACAGAAATATCAATATTATCAGAGA	4683
Db	393	eLysasnspLysglYasprIlethrThraspProthrcIulegInThrhrrIleArgSl	413
OY	4684	ATATTATTAACACTCTTATGCAAAATTAACCTAGAAAATTTAGAGAAATGATTAATTCCT	4743
Db	413	utyTytyrIshIsIeuTYraIaaSnLysLueglIuaSnIueuInglIumetasPrtnPhele	433
OY	4744	GGACACATATGAGCCTGTATGAGACCTTGGGGGACAGAAACAAAAGGGGTGATGACAAA	4803
Db	433	U-----	433
OY	4804	ATAAAAGACAAGACAAAAGATATGTTTGGAAAGTAGGGGTGAGGGGCAACTTGCTCT	4863
Db	433	---	433
OY	4864	AATGACAGAGGCCCTGAGCTTTACACCAACCCTCTGTATTATTAGCAAAAAGAGATAGC	4922
Db	433	---	433
OY	4924	GAGAGGGTGAATTGGAAAGAAAGAGTGAAGTGTAGTTCAGAGTAGGCGCTGCAAGATGTC	4983
Db	433	---	433
OY	4984	ATTCCCTCAACAATAGGCTTAGATGTCCAGTAGATTAACCTCAAGAGCAAGTGCAGG	5043
Db	433	---	433
OY	5044	GAGTGAATGCCCTCAGCAAAACCTCTTAGGGCAGGACGACGAAGTAAGTTGGCCCATCTT	5103
Db	433	---	433
OY	5104	GTAATCAGATAAACAGTTTGCTGTATTATCAAGTAGAGCTCAGTGAATGCTGAGTTGG	5163
Db	433	---	433
OY	5164	TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACCA	5223
Db	434	---AsnthTytrThlrIueProthIueuAnsl	443

QY	5224	GGAGAGACTCAAAATCCCTGGAATATACAGATACAAAGTTCTTAATTTGACAGCATTAATGCA	5283
Db	443	ngluniluvaiIuIsertLeuasnAaGProIleThGlySerclunIleValAlaIlelleAs	463
QY	5284	TAGCCTTACCAACCAAAAAAAGCTCCAGAGCCAGAGGATGTTCACAGCCAAATCTCACGAG	5343
Db	463	nserLeuProThryLysSerProGlyProAspGlyPheThrIaGupPheTyGlnAr	483
QY	5344	GTACAAAGAGAGAGGAGTACTATTCCTTGTGAACCTATTCCAAAAATAGAA---AATGC	5400
Db	483	gTyrlYsGlnIuIeuValIProPheLeuLeuYsLeuPheGlnserThrGluYsGln-G	503
QY	5401	GAATCTCTCCCTACTACTATTTTACAGAGCCAGCATATCTGTATACCAAAACCTTAGCATG	5460
Db	503	lyIleLeuProAsnSerPheTyrgIuaIasertIleIleuIleProLysProGlyIaG	523
QY	5461	ACACAACAAAAAGAGAAATTTTCAGGCCCATATCCCTGATATACATGATGATGAAATTC	5520
Db	523	spThrThLysLysGlnAsnPhaArgProIleSerLeuMetAsnIleAspAlaYsIlel	543
QY	5521	TCAATTAATACGTGGCAACCAAAATCCAGACAGCATCAAAAAGCTTATATCAGCATATC	5580
Db	543	euaSnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG	563
QY	5581	AAAGTTGGCGTATCCCTGGGATGCAAGCGCTGTTCAAAATATGCAATCAATAATGAC	5640
Db	563	lnvalGlyPheIleProGlyMetGlnGlyTyPheAsnIleArgLysSerIleAsnValI	583
QY	5641	GCCATACATTAACAGAACCAATGACAAAAACCATATATATCATATACATACAGAA	5700
Db	583	leglnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspIaGul	603
QY	5701	AGGCTTTGTGTCAAAATTCACAGCCCTTCACGCTAAACAAATTCAGTAAATAGATATCG	5760
Db	603	ysAlaPhaSpLysIleGlnGlnInProPheMetLeuYsThrLeuAsnLysLeuGlyIleA	623
QY	5761	ATGCAATGATATCTCAAAATTAATAGAGCTATTATAC-AAACCCACAGCCATATATAC	5819
Db	623	spGlyThTyThPheLysIleIleArgIaIleTyAspLysProThrAlaSnIleIlel	643
QY	5820	TGAATGGGCAAAACTGGAGACATTCCTTGAGAACGCGCAAGCAAGCAAGAGATGCCCTC	5879
Db	643	euaSnGlyGlnLysLeuGlnIaIaPheProLeuYsThrGlyThArgGlnLysProL	663
QY	5880	TCTCACCACTCTATTCAGATACTATGTGAAGTCTGGCCAGGGCAATCAGCCAAATAGA	5939
Db	663	euserProLeuLeuPheAsnIleValleuGlnValleuAlaIaGAlaIleArgGlnGul	683
QY	5940	AAGAATATAGGGATTTCAAAATAGAAAGAGAGAACTATATGTCTGTGTTCCAGATG	5999
Db	683	ysGlnIleLysGlyIleGlnLeuGlyLysGlnGlnValLysLeuSerLeuPheAlaSpa	703
QY	6000	ACATGTTGTATATTATAGAAACCCATCGTCTACAGCCAAAACTCCTTAAGCTGAATA	6059
Db	703	spMetIleValTyIleuGlnAsnProIleValSerAlaGlnAsnLeuLysLeuIleS	723
QY	6060	GCAACTTGACGAAGTCTCAGAGACAACAATCAATGTGCAAAATTCACAACATCTTAT	6119
Db	723	erAsnPhaSerLysAlaSerGlyTyrlYsIleAsnValGlnLysSerGlnIaIaPheLeuT	743
QY	6120	ACGCCAATATAGACAAACAGAGAGCCAAATCATGAGTCAACTCATTCACAATTGCTA	6179
Db	743	ytThrAsnAsnArgGlnThrGlnSerGlnIleMetSerGlnLeuProPheThrIleAlaS	763
QY	6180	CAAGACAGAAATTAATCTAGGAATACAACTTACAAGGACACGATAGAACTCTTCAAG	6239
Db	763	erLysArgIleLysTyIleuGlyIleGlnLeuThrArgAspValLysAspLeuPheYsG	783
QY	6240	AGAACTACAAACCACTGTCAAGAAATTAAGAGAGACACAACAAATGAAAAACAATTC	6299
Db	783	lAsnTyTyLysProLeuLeuLysGlnIleYsGlnAspThrAsnLysTyPrlYsAsnIleP	803
QY	6300	CATGCTCACAGATAGTAAGAATCAT-----GAATAATGCCATACTGCCCAAGTAATT	6352

```

Db      803  rocySerTrpValGlyArgIleasnIleValIysMetAlaIleLeuProIysValIleT 823
QY      6535  ATATATCAGTGTACCCCATCAAGTACATGACTTCTTCCAGCAATTTGAAAAA 6412
Db      823  yTArgPheAsnAlaIleProIleIysLeuProMetThrPhePheThrGluLeuGluIysT 843
QY      6413  CAACCTTAATTCATATGGAACCAAAAAAGAGCCACAGCCCAAGCAATCTTAAAGC 6472
Db      843  hTThrLeuLysPheIleTrpAsnGln-LysArgAlaIleAlaIleLysAlaIleLeuSer 862
QY      6473  AAAAAGAACAAAGCTGGAGGTATGATGCTACCTGACTTAACATTAATACCTTAAGCTTACA 6532
Db      863  GlnIysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyTYrLysAlaThr 882
QY      6533  GTAACCAAACTGCTAGTACTGTACCAAAAAGATATATAGACCAATGGAAGACA 6592
Db      883  ValThrLysThrAlaTrpIlyTrpIlyGlnAsnArgSpleAspGlnTrpAsnArgThr 902
QY      6593  GAGACCTCAGAAATTAC-ACTCGAATCTACATCCATGATCTTTGACAAACCTGACAAA 6651
Db      903  GluProSerGluIleMetProHisIleTyAsnTyIleuIlePheAspLysProGluLys 922
QY      6652  AACAGCAATGGAAAAAGATTCCTTATTAATAGGTGTGAAAAACTGGCTAGCC 6711
Db      923  AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla 942
QY      6712  ATATGCAAAAAGCTGAAAGTGGATCCCTTCCCTACACTTATACAAAGTAACTCAGA 6771
Db      943  IleGlyArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAsnSerArg 962
QY      6772  TGAATTAAAGCTTAATATATAGACATTAACCAATAAACCCCA-GAAGAAACCTTAGGC 6830
Db      963  TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGly 982
QY      6831  AATACCTTCAAGTATGAGCATGGCAAGACTTCATGATTAAGAACCAAGCAATG 6890
Db      983  IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysTrpProLysAlaMet 1002
QY      6891  GCACAACAAAGCCAAATATAGCAAGTGGATGATGATTAACATATAGACTTCTGACAGCA 6950
Db      1003  AleThrLysAlaLysIleAspLysTrpAspLeuLysLeuLysSerPheCysThrAla 1022
QY      6951  AAAAAAACCTGCTCAGAGTGAACAAGCAACCTCAGCAATGGAGAAAAATTTTGCAT 7010
Db      1023  -LysGluThrThrIleArgAlaAsnArgGlnProThrThrTrpLysIlePheAlaThr 1042
QY      7011  CTATGATCTGACAAAGGCTAATATCCAGAGATCTAGAGAACTTAACAAATTTCAA 7070
Db      1042  rTyISeSerAspLysGlyLeuIleSerArgIleTyAsnGluLeuLysGlnIleTyLys 1062
QY      7071  GAAAAA-----ACAAACCCCGTCAAAATATGGGCAAGATATGAGCAACACTTCAAA 7126
Db      1062  sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys 1082
QY      7127  AGAAGACATTTATGACACCAACAACATATGAAAAAACTCATCATCATTTGGCTTAG 7186
Db      1082  sGluAspIleTyTrpAlaLysLysIleMetLysLysCysSerSerLeuAlaIleArg 1102
QY      7187  AGAAATGCAAAACCAACAGTACATACCATCTCATGCTAGTGAATGGATGATCAC 7246
Db      1102  gGluMetGlnIleLysThrThrMetArgCysHisPheThrProValArgMetAlaIleIle 1122
QY      7247  TAAAAAGTCAAGAAACAAATGCTGAGAGATGTGAGAAATAGCAACACTTTCCA 7306
Db      1122  eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleLysThrLeuLeuHis 1142
QY      7307  CTGTGGTGGGAATGTAATTAATGTTCAACATTTGTGGAAGACAGTGTGAGATTCCTTAA 7366
Db      1142  sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 1162
QY      7367  GGAATCTGAACCAAGAAATATCATTTGACCAAGCAATCCCATCTACTAGTATATACCAA 7426

```

```

Db      1162  gAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProLys 1182
QY      7427  GGAATATTAATCATTTTATTATTAAGACACATGCACACATTTGTTATTCAGACACTGAT 7486
Db      1182  sAspTyLysSerCysCysTyTrLysAspThrCysThrArgMetPheIleAlaIleAlaLeuPh 1202
QY      7487  CACAATAGCAAAAGACTTGGAAACCAACCAATGTCCATCAGTATAGACTGATTAAGAA 7546
Db      1202  eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLys 1222
QY      7547  AACATGCAACATATACACCATGAAATATCTATGACGCCATATAAAG-GATGAGTTCATGTC 7605
Db      1222  sMetTrpHisIleTyThrMetGluTyTrpAlaAlaIleLysAsnAspGluPheMetSer 1242
QY      7606  CTTTGACAGATATGATGATGAAGCGGAACCATCATCTGACAAACCTAACCAAGACA 7665
Db      1242  rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGlu 1262
QY      7666  GAAACCAAAACCCACCATGTTCTCAGCTTGAATGGAGGT 7705
Db      1262  nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 13
ID      000368  PRELIMINARY;  PRT;  1275  AA.
AC      000368;
DT      01-JUL-1997  (TrEMBLrel. 04, Created)
DT      01-JUL-1997  (TrEMBLrel. 04, Last sequence update)
DT      01-JUN-2002  (TrEMBLrel. 21, Last annotation update)
DE      Putative p150.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-97285120; PubMed-9140393;
RA      Sasseman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA      Debernardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H., Jr.;
RT      "Many human LI elements are capable of retrotransposition.";
RL      Nat. Genet. 16:37-43(1997).
DR      EMBL; U93568; AAC51269.1; -.
DR      InterPro; IPR005135; Exo_endo_phos.
DR      InterPro; IPR000477; RVTse.
DR      Pfam; PF03372; Exo_endo_phos; 1.
DR      Pfam; PF00078; Rvtc; 1.
KW      RNA-directed DNA polymerase.
SQ      SEQUENCE 1275 AA; 149078 MW; 3BBC3B2DC2E06B61 CRC64;

Alignment Scores:
Pred. No.: 0
Score: 5163.00
Percent Similarity: 78.83%
Best local Similarity: 73.33%
Query Match: 9.87%
Indels: 173
Gaps: 7

US-10-083-853b-2 (1-29921) x 000368 (1-1275)
QY      3428  ATGGCAGACCAAAATTCACACATATTAATTAACCTTAATGTAATGGGCTAAATTC 3487
Db      1  MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY      3488  CCAATTAAGACACACAGACTGCAAAATTTGGATAAAGAGTCAAGACCCATGCTGCTGT 3547
Db      21  AlaIleLysArgHisArgLeuAlaIleSerTrpIleLysSerLinsProSerValCysCys 40
QY      3548  ATTTCAGAGGCCCATCTCACATGAAAAGACACACATAGGCTCAAAATTAAGGATGGAG 3607
Db      41  IleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
QY      3608  AAGATTTCACAGTAAGTGAACAAAAACAAAAAGAGGGGTGCAATCTAGTCTCT 3667

```



Db	61	LysIleTyrGlnAlaAsnGlyLysGln---	LysLysAlaGlyValAlaIleLeuValSer	79
Qy	3668	GATAAAACAGACTTTAAACCAACAAAGATCAAAAGACAGACAAAGAGCCATTACATAATG		3727
Db	80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet		99
Qy	3728	GTAAGGCATCAATGGAAACAGAGAGCTAACTATCCTAATAATATACATGACCCCAATACA		3787
Db	100	ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr		119
Qy	3788	GGAGCACCCAGATTCTATAAAGCAAGTCTTAGAGACCTCAAAAGAGACTTTTGACTCCAC		3847
Db	120	GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis		139
Qy	3848	ACAATAATAGTGGGAGTCTAAATAAATAATAGACACTTTAAACACCCCACTGCCAATATTA		3907
Db	140	ThrLeuIleMetGly-----	AspPheAsnThrProLeuSerThrLeu	153
Qy	3908	GGCAGATCAATGAGACAGAAAATTAAACAGGATATCCAGAGTTGAACTGAGCTCTGGAC		3967
Db	154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis		173
Qy	3968	CAAGCGACCTAATAGATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTC		4027
Db	174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe		193
Qy	4028	TTCTCAGCATCACATACACTATTTTAAATTTAAATTTGACCATGTAATTTTAACTAAACACTC		4087
Db	194	PheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeuGlySerLysAlaLeu		213
Qy	4088	CTCAGCAATGCAAAAGACAGAAAATCTCAACAAACAGTCTCTCAGACTCAGTGCATC		4147
Db	214	LeuSerLysCysLysArgThrGlnIleIleThrAsnTyrLeuSerAspHisSerAlaIle		233
Qy	4148	TATTTAGAACTCAATTAAGAACCTCACTCAAAATCACACACTACATGAACTGAAC		4207
Db	234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn		253
Qy	4208	AACTGCTCTGTAATGACTACTGGTAAATAAACAATAAGGCAAAAATAAAGATGTC		4267
Db	254	AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaAspIleLysMetPhe		273
Qy	4268	TTTGAACCAATGAGAACAGACACAATGTACAGAACTCTCGGGGCATATTTAAAGCA		4327
Db	274	PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla		293
Qy	4328	GTGTGTAGGGAAATTTATAGACACTAGATGCCTTACAGAGAGAGAGCAAAATATCTAAA		4387
Db	294	ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys		313
Qy	4388	ATAGACACCTTAACATCACAAATTAAGAACTAGAGAAAGAGAAAGCAAAATTTCAAAA		4447
Db	314	IleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGlnGlnThrHisSerLys		333
Qy	4448	GCTAGCAGACAGCAAGAAATTAAGATCAGACGAACTCAAGGAGATAGACACAA		4507
Db	334	AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln		353
Qy	4508	AAAGCCCTTCAATAATCAATCAATCCAGAGCTGTTTTTGAAGAAGATCAGCAAAAT		4567
Db	354	LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle		373
Qy	4568	----AGACCACTAGACAGACTAATAAGAAAGAAAGAGAAAGTCAAGAGATGCAAT		4623
Db	373	eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle		393
Qy	4624	AAAAATGATAAGGGGATATCACCCGATCCCAAGAAATACAAACTATTATCAGAGA		4683
Db	393	eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGlu		413
Qy	4684	ATATTATAACACTCTCTCAATAAATACTAGAAAATCTAGAGAAATGATAAATTCCT		4743
Db	413	uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe		433
Qy	4744	GGACACATATGTAGCTGTATGGACCTTTGGGGGACAGACAAAGGGGGTGAATGCAGAA		4803
Db	433	uAsp-----		434
Qy	4804	ATAAAGACAAAGACAAAGAGATATGTTTGGAAAGTAGGGTTCAGGGGGCAACTTGCCTCT		4863
Db	434	-----		434
Qy	4864	AATGGACAAGGGCCCTGAGCTTTAGACCACCTCTGTATTTATTATTAGCAAAAGAGATAGC		4923
Db	434	-----		434
Qy	4924	GAGAGGTGATTTGGAAAGAGAGGTACAGCTGTTAGTCCAGAGTAGGCTTGAAGACTGC		4983
Db	434	-----		434
Qy	4984	ATTCTCTCAAAATAGAGCTCTAGATGTCCAGTAGATAAACCCTCAAGGAGCCAGTGCAGG		5043
Db	434	-----		434
Qy	5044	GAGTGATGGCCCTCAGCAAAACCTTCTAGGCGCAGCAGCAAGTAAGTTTGCCACATCT		5103
Db	434	-----		434
Qy	5104	GTATTCAGATAAACAGCTTTGCTGTTGATCAAGTAGCTCCAGTCCAGTGCAGTTGG		5163
Db	434	-----		434
Qy	5164	TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAGACTAAACCA		5223
Db	435	-----	ThrTyrThrLeuProArgLeuAsnGlu	443
Qy	5224	GGAGAGACTCAATCCCTGAATATACCAAGTCTCAAAATTTCAAAATTTCAAGCAAGTGA		5283
Db	443	nGluGluValGluSerLeuAsnArgArgIleThrGlySerGluIleValAlaIleLeuAs		463
Qy	5284	TAGCTACCAACCAACCAAAAGTCCAGGACCAAGGATTCACAGCCAAATTTCTACACAG		5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg		483
Qy	5344	GTACAAAGAGAAGCTGTGTACTATTCCTCTGAAACTATTCACAAAATAAGAA---AATGG		5400
Db	483	gTyrMetGluGluValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G		503
Qy	5401	GAATCCCTCCCTAACTCACTTTTACGAGGCCAGCATCTCGATACCAAAACCTAGCAGTG		5460
Db	503	lyleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArg		523
Qy	5461	ACACAACAAAGAGGAAATTTTCAGGCCCATATCCCTGTGATGATGATGTGAAAATCC		5520
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL		543
Qy	5521	TCAATAAATATCTGGCAACCAATTCAGCAGACATCAAAAGCTTATCTACCATCATC		5580
Db	543	euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisasp		563
Qy	5581	AAGTGGCGTCATCCCTGGCATGCAAGCTGTTCAAAATATGCAAAATCAATAATCTAG		5640
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleasnValI		583
Qy	5641	GCCATCACATAAAGACAGAAATTCACAAAACCATGATTTCTCAATAGATGCAGAAA		5700
Db	583	leGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleSerIleAspAlaGluL		603
Qy	5701	AGGCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATTCAGTAAGTACTAGGTATCG		5760
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA		623
Qy	5761	ATGCAATGTATCTCAAAATATAGACTATTTATATAC-AAACCCACAGCAAAATATCATAC		5819
Db	623	spGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL		643





```
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93564; AAC51263.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; Rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149201 MW; 23D516D6B4358F28 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5150.00 Matches: 1053
Percent Similarity: 78.69% Conservative: 77
Best Local Similarity: 73.33% Mismatches: 140
Query Match: 9.84% Indels: 173
DB: 4 Gaps: 7

US-10-083-853B-2 (1-29921) x 000362 (1-1275)
QY 3428 ATGGCAGGATCAATTCACACATAATAATTAACCTTAATGTAATGGCTAAATTC 3487
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY 3488 CCAATTAAAGACACAGACTGGCAATTTGGATAAAGAGTCAAGACCCATCAGTGTCTGT 3547
DB 21 AlaIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTGAGGAGGCCATCTCACATGAAGAAGACACACATAGGCTCAAAATAAAGGATGAGG 3607
DB 41 IleGlnIleThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
QY 3608 AAGATTACCAAGTAAATGAAACAAACAAAAAGAGCGGTGCAATCTAGTCTCT 3667
DB 61 LysIleTrpGlnAlaAsnGlyLysGln--LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATAAAGACAGCTTTAAACCAACAAAGATCAAAAGAGACAAAGGCCATTCATCAATG 3727
DB 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet 99
QY 3728 GTAAGGCATCAATGGACAGAGAGAGCTAATCTTAAATATACATGACCCCAATACA 3787
DB 100 ValLysIleSerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr 119
QY 3788 GGAGCACCAGATTCATAAGCAAGTCTTAGAGACCTACAAAGAGACTTTGACTCCAC 3847
DB 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
QY 3848 ACAATAATAGTGGAGTCTAAATAATAATAGACACTTTAACACCCCACTGCCAATATTA 3907
DB 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3908 GCAGATCAATGACAGAGAAATTAACAGGATATCAGGAGTGAAGTGAAGTCTGGAC 3967
DB 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
QY 3968 CAAGCGGACCTAATAGATATCTACAGAACCTCCCAACCCCAACAAATCAACAGATATACACTC 4027
DB 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193
QY 4028 TTCTGACATCAGATACACTATTTTAAATGACCATGTAATTTAAGTAAACACTC 4087
DB 194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213
QY 4088 CTCAGCAATGCAAAAGACAGAAATCTTAACAAACAGTCTCTCAGACTCAGTGCATC 4147
DB 214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
QY 4148 TATTAGAACTCAGAAATTAAGAACTCCTCAAAATCACACAACTACATGAACTGAAC 4207
DB 234 LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn 253
QY 4208 AACCTGCTCTGAATGACTACTGGGTAATAATACAAATGAAGGCAAAATTAAGATGTC 4267
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
QY 4268 TTTGAAACCAATGAGACAAAGACACAAATCTACAGAAATCTCTGGGCGCATATTTAAAGCA 4327
DB PheGluThrAsnGlnAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293
QY 4328 GTGTGTAGAGGAAATTTATAGCACTAGATGCGCTACAAAGAGAAAGACAGAAATCTAAA 4387
DB 294 ValCysArgGlyLysLeuIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
QY 4388 ATAGACACCTTAACATCACATAATTAAGAACTAGAGAAAGAGAGCAAAACAATTCAAA 4447
DB 314 IleAspThrLeuThrSerGlnLeuLysGluLeuLysGlnGluInThrHisSerLys 333
QY 4448 GCTAGCAGAGACAAAGAAATAACTAAGATCAGACAGCAACTGAGGAGATAGAGACAAA 4507
DB 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
QY 4508 AAAAGCCCTTCAAAATAAATCAATGAATCCAGGAGCTGTTTTTTGAAAAGATCAGCAAAAT 4567
DB 354 LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle 373
QY 4568 ----AGACCACTAGACAGACTAATAAGAAAGAAAGAGAGAAAGATCAAGAGATCAAT 4623
DB 373 eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle 393
QY 4624 AAAAATGATAAAGGGATATACACCCGATCCACAGAAATCAACAATATATTCACAGA 4683
DB 393 eLysAsnAspLysGlyAspIleThrAspProSerGluIleGlnThrIleArgGln 413
QY 4684 ATATTATAACACCTCTATGCAATTAACATAGAAAATCTAGAAAATGATGATAAATTCCT 4743
DB 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe 433
QY 4744 GGACACATATGTAGCCTGTATGGACCTTGGGGGACAGACAAAGGGGGTGAATGCAGAA 4803
DB 433 uAsp----- 434
QY 4804 ATAAAGACAAAGACAAAAGAGATATGTTTGAAGTAGGGCTCAGGGGGCAACTTGCCTCT 4863
DB 434 ----- 434
QY 4864 AATGGACAGGGCCCTGAGCTTTACACACCCCTCTGTATTTATTAGGCAAAAGAGATAGC 4923
DB 434 ----- 434
QY 4924 GAGAGGTGAGTTGGAAGAAGAGGTGAGCTTTAGGTCCAGAGTAGGCCCTGCAAGACTGC 4983
DB 434 ----- 434
QY 4984 ATTCCTCAAAACAATAGCTCTAGATGTCCAGTAGATAACCTCAAGGAGGAGTGCAGG 5043
DB 434 ----- 434
QY 5044 GAGTGTAGGCCCTCAGCAAACTTCTAGGCGAGGCACAGAAAGTAAGTTGCCACATCTCT 5103
DB 434 ----- 434
QY 5104 GTATTACAGATAAACAGTTTGTGTTGATCAAGTAGCCCTCCAGTGAATGCTGAGTTGG 5163
DB 434 ----- 434
QY 5164 TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTTAAACCA 5223
DB 435 -----ThyTyrThrLeuProArgLeuAsnGln 443
QY 5224 GGAGAGAGTCAATCCCTGAATATATACCAAGTTCTAAAATTTGAAGCAGTAAATGTA 5283
DB 443 nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleLeuAs 463
QY 5284 TAGCCTACCAACCAAAAGAGTCCAGGACAGCGGATTCACAGCCAAATTTACACAGAG 5343
DB nSerLeuProThrLysLysSerLeuGlyProAspArgPheThrAlaGluPheTyrGlnArg 483
```

QY	5344	GTACAAGAGAACTGGTACTATTCTCTGTAACATCTTCCAAAAAATAGAA---AATGG	5400
Db	483	gTyLysGluGluLeuValProPheLeuLysLeuPheGlnSerIleGluLysGlu-G	503
QY	5401	GAATCTCCCTAACTATTACGAGGCGAGCATCTCTGATACCAAAACCTAGCAGTG	5460
Db	503	lylleLeuProAsnSerPheTyrgluAlaSerllelleLeuIleProLysProGlyArgA	523
QY	5461	ACACAACAAAAGAGAAATTTCCAGGCCATATCCCTGATGAACTGATGCAAAATCC	5520
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543
QY	5521	TCAATAAATACGTGCAACCAATCCAGCAGCACATCAAAAAAGCTTATCTACCATGATC	5580
Db	543	euAsnLysIleLeuAlaAsnArgIleGlnProIleLysLysLeuIleHisAspG	563
QY	5581	AGTTGGCTCATCCCTGGATCCAGGCTGGTTCAAAATATGCAATCAATAATGTAG	5640
Db	563	InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI	583
QY	5641	GCCATCACATAAACAACAACATGACAAAAACACATGATTATCTCAATAGATGCAGAA	5700
Db	583	leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL	603
QY	5701	AGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAAAAATTCAGTAAACTAGGTATCG	5760
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623
QY	5761	ATGGATGTATCTCAAAATTAATAAGAGCTATTATTATAC-ANACCCAGCAATATCATAC	5819
Db	623	spGlyThrTyrrPheLysIleIleArgAlaIleTyrrAspLysProThrAlaAsnIleL	643
QY	5820	TGAATGGCAAAACCTGGAAGCATTCCTTTGAGAACTGGCACAAGAGATGCCCTC	5879
Db	643	euAsnGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663
QY	5880	TCTCACCCTCTATTCAAGACTATTGGAAGTTCTGGCCAGGGAATCAGGCAATAGA	5939
Db	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683
QY	5940	AGAAATAAGGGTATTCAATAGAAAGAGAGGAAGTATATGCTCTGTTGCGAGATG	5999
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703
QY	6000	ACATGTTTGTATATTTAGAAAAACCCATCGTCTCAGGCCAAAAACCTCTTAAGCTGATAA	6059
Db	703	spMetIleValTyrrLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleS	723
QY	6060	GCACCTTACGAAAGTCTCAGGACACAAATCAATGTGCAAAATCACAAAGCATTTAT	6119
Db	723	erAsnPheSerLysValSerGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeuT	743
QY	6120	AGCCCAATATAGACAAACAGACCAATCATGATGATCACTCATTCACAAATGCTA	6179
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyLeuProPheThrIleAlaAs	763
QY	6180	CAAGAGAATAAATACCTAGGAATACAACTTACAGGGACACGATAGGAACCTCTCAAGG	6239
Db	763	erLysArgIleLysTyrrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783
QY	6240	AGAATCAAAACCTGATCAAGGAATAAGAGAGAGACACAAACAAATGGAAAAACATTC	6299
Db	783	luAsnTyrrLysProLeuLysGluIleLysGluAspThrAsnLysTyrrLysAsnIleP	803
QY	6300	CATGCTCAGACATA-----GTAAGAAATCAATAAATGCC-ATACTGCCCAAGTAATT	6352
Db	803	roCysSerTrpValGlyArgIleAsnIleMetLysMetAlaIleLeuProLysValIleT	823
QY	6353	ATAGATTCAGTCTACCCCATCAAGCTACCATCTTCTTCTTCTCAGAAATTTGGAAAAA	6412
Db	823	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843
QY	6413	CAACTTTAAATTTATATGTAACCAAAAAAGAGCCCAAGCAACATCTTAAAGC	6472
Db	843	hrThrLysLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862
QY	6473	AAAAAGAACAAAGCTGGAGTATCATCTGCTACCTGACTTAAACTATACATATAGGCTACA	6532
Db	863	GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrrLysAlaThr	882
QY	6533	GTAACCAAACTGCATGGTACTGACCAAAAAACAGATATATAGCAAAATGGACAGAAC	6592
Db	883	ValThrLysThrAlaTrpTyrrTyrrGlnAsnArgAspIleAspGlnTrpAsnArgThr	902
QY	6593	GAGACCTCAGAAATTAAC-ACTGCAATCTACATCCATCTGATCTTTGACAAACCTGCACAA	6651
Db	903	GluProSerGluIleMetProHisIleTyrrAsnTyrrLeuIlePheAspLysProGluLys	922
QY	6652	ACACAGCAATGAAAGAGATTCCTTATTAATGCTTGGAAAACTGGCTAGGC	6711
Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
QY	6712	ATATGCAAGAAAGCTGAACTGGATCCCTTCTTACACCTTATACAAAAGTTAACTCAAGA	6771
Db	943	IleCysArgLysLysLeuLysLeuAspProPheLeuThrProTyrrThrLysIleAsnSerArg	962
QY	6772	TGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCCA-GAAGAAACCTAGGC	6830
Db	963	TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnArgGly	982
QY	6831	AATACCATTCAGATATGACATGGCAAGACTTCATCAGTAAACACCAAAACCAATG	6890
Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
QY	6891	GCAACAAAAGCCAAATAGACAAGTGGGATCTGATTAAACTATAGAGCTTCGCACAGCA	6950
Db	1003	AlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
QY	6951	AAAAAAACTGTCATCAGAGTGAACAAGCAACCTACAGAATGGGAAAAATTTTGCAT	7010
Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh	1042
QY	7011	CTATCGATCTGACAAGGCTAATATCCAGAGATCTACGAAGACTTAAACAAATTTACAA	7070
Db	1042	rTyrrSerSerAspLysGlyLeuIleSerArgIleTyrrAsnGluLysGlnIleTyrrLy	1062
QY	7071	GAAGAAA---AAACACCCCGTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAA	7126
Db	1062	sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLy	1082
QY	7127	AGAAGACATTTATGAGCCACACACATATGAAAAAACCTCATCATCTTGGTCGTAG	7186
Db	1082	sGluAspIleTyrrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr	1102
QY	7187	AGAAATGCAAAACAAACACACAGTGCATACCATCTCATGCTAGTATAGATGGTGATCAC	7246
Db	1102	gGluMetGlnIleLysThrThrMetArgTyrrHisLeuThrProValArgMetAlaIleL	1122
QY	7247	TAAAAAGTCAGGAACAAACAAATGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTCCA	7306
Db	1122	eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuH	1142
QY	7307	CTGTTGGTGGGAATGTAATTAGTTCAACCATTTGGGAAGACAGCTGTGGAGATTCTTAA	7366
Db	1142	sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr	1162
QY	7367	GGATCTAGACACAGAAATATCATTTGACCCAGCAATCCCATCTAGTATATATACCCAAA	7426
Db	1162	gasPheGluLeuGluIleProPheTyrrProAlaIleProLeuLeuGlyIleTyrrProLy	1182
QY	7427	GGAATATAAATCATTTATTATAAGACACATATGACACATATGTTTATTGACAGCATGT	7486
Db	1182	sAspTyrrLysSerCysTyrrLysAspThrCysThrArgMetPheIleGluAlaLeuPh	1202
QY	7487	CACAAATAGCAAGACTTGGACCAACCAACCAATGTCCATCATGATGATGATGATAAGNA	7546

Db	1202	eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLy	1222
QY	7547	AACATGGCACATATACACCACTAAATACTATGAGCCATATAAAG-CATGAGTTCATGTC	7605
Db	1222	smetTrpHisIleThrMetGluTrpTrpAlaAlaIleLysAsnAspGluPheIleSe	1242
QY	7606	CTTTGCAGAGATATGATGAAGCTGGAACCATCATCTCAGCAAACTAACACAAACA	7665
Db	1242	rPheValGlyThrTrpMetLysLeuGluThrIleLeuSerLysLeuSerGlnGluG	1262
QY	7666	GAAACCAACACACATGTTCTCAGTTCAGTGAAGTGGGAGT	7705
Db	1262	nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn	1275
RESULT 15			
O00372		PRELIMINARY;	PRT; 1275 AA.
ID	O00372		
AC	O00372;		
DT	01-JUL-1997	(TrEMBLrel. 04, Created)	
DT	01-JUL-1997	(TrEMBLrel. 04, Last sequence update)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)	
DE	Putative p150.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97285120; PubMed=9140393;		
RA	Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,		
RA	DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;		
RT	"Many human LI elements are capable of retrotransposition.;"		
RL	Nat. Genet. 16:37-43(1997).		
DR	EMBL: U93570; AAC51273.1; -		
DR	InterPro: IPR005135; Exo_endo_phos.		
DR	InterPro: IPR000477; RVTse.		
DR	Pfam: PF03372; Exo_endo_phos; 1.		
DR	Pfam: PF00078; rvt; 1.		
KW	RNA-directed DNA polymerase.		
SQ	SEQUENCE 1275 AA; 149111 MW; 4711B3BC22F7674E CRC64;		
Alignment Scores:			
Pred. No.:	0	Length:	1275
Score:	5133.00	Matches:	1045
Percent Similarity:	78.55%	Conservative:	83
Best Local Similarity:	72.77%	Mismatches:	142
Query Match:	9.81%	Indels:	173
DB:	4	Gaps:	7
US-10-083-853B-2 (1-29921) x O00372 (1-1275)			
QY	3428	ATGGCAGGATCAATTCACACATAATAATTAACCTTAATTAATGGCTAAATTC	3487
Db	1	MethrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer	20
QY	3488	CCAATTAAGACACACAGACTGGCAATTTGGATAAAGACTCAAGACCCATCAGTGTGCTGT	3547
Db	21	AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys	40
QY	3548	ATTGAGAGGCCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGGATGGAGG	3607
Db	41	IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTyrArg	60
QY	3608	AAGATTACCAAGTAATGGAACAAACAAAAAAGACGAGGGTTCGAATCCTAGTCTCT	3667
Db	61	LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer	79
QY	3668	GATAAAACAGACTTTAAACCAACAAAGATCAAAAGACACAAAGAGCCCATACATAATG	3727
Db	80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet	99
QY	3728	GTAAGGCATCAATGGAAACAAGAGAGCTAACTATCCTAAATATACATGACCCCAATACA	3787

Db	100	VaIysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr	119
QY	3788	GGAGCAGCCAGATTTCATAAGCAGTTCCTTAGAGACCTACAAAGAGACTTTCAGCTCCAC	3847
Db	120	GlyAlaProArgPheIleLysGlnAspLeuSerAspLeuArgAspLeuAspSerHis	139
QY	3848	ACAATAATAGTGGGAGTCTAAATAATAATAGACACTTTAACACCCCACTGCCAATATTA	3907
Db	140	ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu	153
QY	3908	GGCAGATCAATGACACAGAAAAATTAAACAGATATCCAGAGTTGAACTGAGCTCTGGAC	3967
Db	154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis	173
QY	3968	CAAGCGGACCTAATAGATATCTACAGAACTCCCAACCCCACTAACTAACAGATATATAC	4027
Db	174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe	193
QY	4028	TTCTCAGCATCACATTACACTATTTTAAATTTGACCATGTAATTTTAAAGTAAACACTC	4087
Db	194	PheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeuGlySerLysAlaLeu	213
QY	4088	CTCAGCAATGCAAAAGACAGAAATCCTTAACAAACAGTCTCTCAGACTACAGTGCATC	4147
Db	214	LeuSerLysCysLysArgThrGluIleThrAsnTyrLeuSerAspHisSerAlaIle	233
QY	4148	TATTAGAATCTAGAACTACAACTCCTCAAAATCACAACTACATGAACTGAACTGAAC	4207
Db	234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn	253
QY	4208	AACCTGCTCTGATGACTACTGGTAAATAACAAAAATGAAGCGCAAAATAAAGATGTC	4267
Db	254	AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe	273
QY	4268	TTTGAACCAATGAGACAAAGACACAATGTACACAGAACTCTCTGGGCGCATATTTAAAGCA	4327
Db	274	PheGluThrAsnLysAsnLysAspThrThrTyrGlnAsnLeuTyrAspThrPheLysAla	293
QY	4328	GTGTGTAGAGGGAATTTATAGCACTAGATGCTTACAGAGAGAAAGAGAGAAATATCAAA	4387
Db	294	VaIysArgGlyLysPheThrAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys	313
QY	4388	ATAGACACCTTAACATCACAAATTAAGAACTAGAGAAAGAGAGAAAGAGAAATTCAAA	4447
Db	314	IleAspThrLeuThrSerGlnLeuLysGluLeuLysGlnGlnGlnThrHisSerLys	333
QY	4448	GCTAGCAGACAGCAAGAAATACTAAGATCAGACGACAGACTGAAGGAGATAGACACAA	4507
Db	334	AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln	353
QY	4508	AAAGCCCTTCAAAATAATCAATCAATCCAGGAGCTGTTTTTCAAAAGATCACGAAAT	4567
Db	354	LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysIle	373
QY	4568	----AGACCACTAGACAGACTTAATAAGAAAGAGAGAGAAAGATCAAGAGATGCAAT	4623
Db	373	eAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIle	393
QY	4624	AAAAATGATAAGGGGATATCACCCGATCCACAGAAATACAAACTATTATTCAGAGA	4683
Db	393	eLysAsnAspLysGlyAspIleThrAspProThrGluIleGlnThrThrIleArgG	413
QY	4684	ATATTATAACACCTCTATCAATTAACCTAGAGAAATCTAGAGAAATGATAATTCCT	4743
Db	413	uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspLysPheLe	433
QY	4744	GGACACATATGTAGCCTGTATGGACCTTGGGGGACAGAAACAAAGGGGGTGAATGCAGA	4803
Db	433	u-----	433
QY	4804	ATAAAGACAAAGACAAAGAGATGTTTGGAGTAGGGGTGAGGGGGCAACTTGCCTCT	4863

Db	433	-----	433
Qy	4864	AATGGACAAGGCCCTGAGCTTTACACCACCCTCTGTATTATTAGGCAAAAGAGATAGC	4923
Db	433	-----	433
Qy	4924	GAGAGGTGAGTTGGAGAAGAGAGGTACAGTGTAGTGTCCAGAGTAGGCGCTGCAAGACTGC	4983
Db	433	-----	433
Qy	4984	ATTCTCTCAACAATAGAGCTCTAGATGTCCAGTAGATAACCTCAAGGAGCGCAGTGCCAGG	5043
Db	433	-----	433
Qy	5044	GAGTGTGGCCCTCAGCAAAACCTTCTAGGCGAGGCACAGAAGTAAGTTTGCCACATCTCT	5103
Db	433	-----	433
Qy	5104	GTATTCAGGATAACAGTTTGTGTTGTGATCAAGTAGCTCCAGTGGGAATGCTGAGTTGG	5163
Db	433	-----	433
Qy	5164	TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACCA	5223
Db	434	-----	434
Qy	5224	GGAAGAAGTCAAAATCCCTGAATATACCAAGTCAACAAGTTCTAAATTTGAAGCAGTAATTGA	5283
Db	443	nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValAlaIleIleAs	463
Qy	5284	TAGCCTACCAACCAAAAAGTCCAGGACGAGGNTTCACAGCCAAATTCACAGAG	5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAspGluPheTyrglnAr	483
Qy	5344	GTACAAAGAGAAGCTGTACTATTCTCTGAACTATTCCAAAAATAGAA--AATGG	5400
Db	483	gTyrlsGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleIleLysGlu-G	503
Qy	5401	GAATCTCCCTAACTCATTTTACGAGCCAGCATCTCTGTATACCAAACTAGCAGTG	5460
Db	503	lytleLeuProAsnSerPheTyrgluAlaSerIleIleLeuIleProLysProGlyArgA	523
Qy	5461	ACACAACAAAGAGGAAATTCAGGCCCATATCCCTGATGACATGATGTAATCC	5520
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543
Qy	5521	TCATAAATACTGCGCAAAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580
Db	543	euAsnLysIleLeuAlaAsnGlnIleGlnHisIleLysLysLeuIleHisAspG	563
Qy	5581	AGTTGGGTATCCCTGGGATGCAAGGCTGGTTCAAAATATATGCAAAATCAATAAATGTAG	5640
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheAsnMetCysLysSerIleAsnValI	583
Qy	5641	GCCATCACATAAACAGAACCAATGACAAAACCCACATGATTCTCAATAGATGCGAAA	5700
Db	583	leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL	603
Qy	5701	AGGCTTTGTCAAATTCACAGCCCTTCATGTGTAAATAATTCAGTAAACTAGGTATCG	5760
Db	603	ysAlaLeuAspLysIleGlnInProPheValLeuLysThrLeuAsnLysLeuGlyIleA	623
Qy	5761	ATGGAATGTATCTCAAAATAAAGAGCTATTATATAC-AAACCCAGGCAATATATCATAC	5819
Db	623	spGlyThrTyrlsPheLysIleMetArgAlaIleTyrlsAspLysProThrAlaAsnIleIleL	643
Qy	5820	TGAATGGCAAAACCTGGAGGATTCCTTTGAGAACTGGCAGACAGAGATGCCCTC	5879
Db	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663
Qy	5880	TCTCACCTCTCTTCAAGACTACTATTGGAAGTTCTGGCCAGGGCAATCAGCAGATAGA	5939
Db	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683
Qy	5940	AAGAAATAAAGGGTATTCAAAATAGAAAGAGAGAGATCATATTGCTCTGTTTCGAGATG	5999
Db	683	ysGluIleLysGlyIleGlnLysGluGluValLysLeuSerLeuPheAlaAspA	703
Qy	6000	ACATGTTTGTATATTAGAAAACCCCATCTCTCAGGCCCAAAACTCCTTAAGCTGATAA	6059
Db	703	spMetIleValTyrlsLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleL	723
Qy	6060	GCAACTTCACAAAGTCTCAGGACACAAATCAATGTGCAAAATCAACAGCACTTCAT	6119
Db	723	erAsnPheSerLysValSerGlyTyrlsIleAsnValGlnLysSerGlnAlaPheLeuT	743
Qy	6120	ACGCCAATAATAGACAAACAGAGCCAAATCATGAGTGAACCTCTCATTCACAAATTCCTA	6179
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlaL	763
Qy	6180	CAAGAGATAAATACCTTAGGAATACAACTTACAGGGACACAGTAGGAATCTTCAAG	6239
Db	763	erLysArgIleLysTyrlsLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783
Qy	6240	AGAACTACAAACCACTGATCAAGGAATAAGAGAGGACACAAACAATAAGAAAACATTC	6299
Db	783	luAsnTyrlsProLeuLeuLysGluIleLysGluAspThrAsnLysIleAsnIleP	803
Qy	6300	CATGCTCACAGATAGTAAGAAATCAT-----GAAAATGCCATCTGCCCAAGTAAAT	6352
Db	803	rocysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823
Qy	6353	ATGATTCAGTGTACCCCATCAAGCTACCATGACTTCTTTCACAGAAATGGAAGAAA	6412
Db	823	yrArgPheAsnAlaIleProIleAsnLeuProMetThrPhePheThrGluLeuGluLysT	843
Qy	6413	CAACTTTAAATTTCAATATGGAACCAAAAAGAGCCACAGAGCCAGACAATCTTAAGC	6472
Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862
Qy	6473	AAAAAGACAAAGCTGGAGGTATCATGCTACCTGACTTAAACTATACTATAAGGCTACA	6532
Db	863	GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyrlsAlaThr	882
Qy	6533	GTACCAAAACTGCATGGTACTGGTACCAAAACAGATATATACCAATGGAACAGACA	6592
Db	883	ValThrLysThrValTrpTyrlsTrpTyrlsGlnAsnArgAspIleAsnTrpLeuAla	902
Qy	6593	GAGACCTCAGAATATAC-ACTGCAATCTACATCCATCTGATCTTTCACAAAACCTGACAA	6651
Db	903	GluProSerGluIleMetProHisIleTyrlsAsnTyrlsMetIlePheAspLysProGluLys	922
Qy	6652	ACAACGAATGGAAGGATTCCTTATTTAATAATGGTGTGGTGGAAAACTGGCTAGCC	6711
Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
Qy	6712	ATATGCAAGAACTGAACTGCTCCCTTCCCTTACACCTTATACAAAGCTTACTCAAGA	6771
Db	943	IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrlsIleAsnSerArg	962
Qy	6772	TGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCA-CAAGAAAACCTAGGC	6830
Db	963	TrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGly	982
Qy	6831	ATATCAATTCAGATATGACATGGCAAGACTTTCATGACTATAAACACCAAAAGCAATG	6890
Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
Qy	6891	GCAACAAAACCAAAATACACAAGTGGGATCTGATTAAACTATAGAGCTTCGCACAGCA	6950
Db	1003	AlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
Qy	6951	AAAAAAACTGTCTCAGAGTGAACCAAGCAACCTACAGAAATGGGAAAAATTTTCAAT	7010
Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh	1042

```
QY 7011 CTATCGATCTGACAAAGGCTTAATATCCAGAGATCTACGAGAACTTAACAAATTTACAA 7070
Db      |||||
1042 rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLy 1062
QY 7071 GAAAAA--AACAAACCCGCTAAAATATGCGCAAGGATATGAGACACACTTCTCAA 7126
Db      |||||
1062 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLy 1082
QY 7127 AGAAGACATTTATGCGCAACAAACATATGAAAAACCTCATCATCTGCTGCTGCTAG 7186
Db      |||||
1082 sGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr 1102
QY 7187 AGAAATGCAAAACAAACACAGTGCATACATCTCATCTAGTGTAGATGTGTATCAC 7246
Db      |||||
1102 gGluMetGlnIleLysThrThrMetArgTyrHisHisThrProValArgMetAlaIleIl 1122
QY 7247 TAAAAAGTCAGGAAACAAACAAATCTGCGAGAGGATGTGGAGAAATAGGACACTTTTCCA 7306
Db      |||||
1122 eLysLysSerGlyAsnAsnArgCysTrpArgCysGlyGluIleGlyThrLeuLeuH1 1142
QY 7307 CTGTTGGTGGGAATGTAATTAAGTCAACCATTTGTGAACACAGTGTGGAGATTCCTTAA 7366
Db      |||||
1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGATCTAGAACACAGAAATATCATTTGACCCAGCAATCCCATTTACTGAGTATATACCCAAA 7426
Db      |||||
1162 gaspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProLy 1182
QY 7427 GGAATATATATCATTTCTATATATAAGACACATGCACACATATGTTTATGCCACACTGAT 7486
Db      |||||
1182 sasptYrLysSerCysCysTrpLysAspThrCysThrArgMetPheIleAlaLeuPh 1202
QY 7487 CACAATAGCAAGACTTGGAAACCAACCAATCTCCATCAGTCATGATAGCTGGATAAGAA 7546
Db      |||||
1202 eThrIleAlaLysThrTrpLysGlnProLysCysProThrMetIleAspTrpIleLysLy 1222
QY 7547 AACATGGCACATATACACCATGAAATACTATGCGCCATAAAAAAG-GATGAGTTCATGTC 7605
Db      |||||
1222 sMetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGAGCTGGAACCATCTCTCAGCAAACTACACAAAGAACA 7665
Db      |||||
1242 rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGluGl 1262
QY 7666 GAAAAACCAACACACATGTTCTCACTTGTAACTGGGAGT 7705
Db      |||||
1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275
```

Search completed: July 4, 2003, 20:41:33  
Job time : 1356 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 4, 2003, 19:01:07 ; Search time 186.5 Seconds  
(without alignments)  
13308.453 Million cell updates/sec

US-10-083-853B-2  
Defect score: 53328  
Sequence: 1 gtatatgaagaagctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO\_SPOOL/US10083853/runat\_03072003\_093611\_8078/app\_query.fasta\_1.30087  
-DB=SwissProt\_40 -OFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pt -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10083853 -CGN=1.1.311 -runat\_03072003\_093611\_8078 -NCPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	5139	9.8	1259	1	LN1_HUMAN
2	3658.5	7.0	1260	1	LN1_NYCCO
3	3548.5	6.8	1300	1	POL2_MOUSE
C 4	2145	4.1	1259	1	LN1_HUMAN
C 5	1479.5	2.8	1260	1	LN1_NYCCO
C 6	1450.5	2.8	1300	1	POL2_MOUSE
C 7	443	0.8	1308	1	YTX2_XENLA
C 8	412	0.8	593	1	ALU7_HUMAN
9	393	0.8	593	1	ALU7_HUMAN
10	392	0.7	593	1	ALU6_HUMAN
11	387	0.7	379	1	YOL1_MOUSE
C 12	385	0.7	591	1	ALU6_HUMAN
C 13	381.5	0.7	591	1	ALU8_HUMAN
C 14	358	0.7	585	1	ALU5_HUMAN
C 15	348.5	0.7	585	1	ALU5_HUMAN
C 16	347	0.7	591	1	ALU1_HUMAN
C 17	345.5	0.7	591	1	ALU8_HUMAN
C 18	334	0.6	587	1	ALU2_HUMAN

C 19	328	0.6	587	1	ALU3_HUMAN	P39190 homo sapien
C 20	320	0.6	587	1	ALU2_HUMAN	P39189 homo sapien
C 21	316.5	0.6	591	1	ALU1_HUMAN	P39188 homo sapien
C 22	308	0.6	587	1	ALU3_HUMAN	P39190 homo sapien
C 23	290	0.6	603	1	ALU4_HUMAN	P39191 homo sapien
C 24	284	0.5	603	1	ALU4_HUMAN	P39191 homo sapien
C 25	246.5	0.5	317	1	TLN_MOUSE	P17408 mus musculus
C 26	241.5	0.5	1025	1	PO21_NASVI	Q03278 nasonia vit
C 27	235	0.4	916	1	RTJK_DROME	P21328 drosophila
C 28	233	0.4	317	1	TLN_MOUSE	P17408 mus musculus
C 29	223	0.4	1057	1	POLR_DROME	P16423 drosophila
C 30	221	0.4	418	1	YY1_HUMAN	P49646 homo sapien
C 31	214.5	0.4	1222	1	YMH5_CAEEL	P34472 caenorhabdi
C 32	211	0.4	711	1	PO22_PORJA	Q03274 popillia ja
C 33	209.5	0.4	1790	1	USO1_YEAST	P25386 saccharomyc
C 34	201	0.4	2230	1	GOG4_HUMAN	Q13439 homo sapien
C 35	199.5	0.4	1957	1	YD86_SCHPO	Q10411 schizosacch
C 36	198.5	0.4	2663	1	GENE_HUMAN	Q02224 homo sapien
C 37	194	0.4	944	1	NUF1_YEAST	P32380 saccharomyc
C 38	194.5	0.4	1805	1	HMW2_MYCGE	P47460 mycoplasma
C 39	193.5	0.4	1875	1	MLP1_YEAST	Q02455 saccharomyc
C 40	193	0.4	2116	1	MY52_DICDI	P08799 dictyosteli
C 41	186	0.4	2869	1	RBP1_PLAVB	Q00798 plasmodium
C 42	182.5	0.3	1251	1	RBP2_PLAVB	Q00799 plasmodium
C 43	181.5	0.3	1005	1	RA50_METJA	Q58718 methanococc
C 44	179	0.3	1312	1	RA50_YEAST	P12753 saccharomyc
C 45	178	0.3	1453	1	Y373_BOVIN	Q9TU23 bos taurus

ALIGNMENTS

RESULT 1  
LN1\_HUMAN  
ID LN1\_HUMAN STANDARD; PRT: 1259 AA.  
AC P08547;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-AUG-1988 (Rel. 08, Last annotation update)  
DE LINE-1 reverse transcriptase homolog.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=86230917; PubMed=2423883;  
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;  
RT "L1 family of repetitive DNA sequences in primates may be derived  
RT from a sequence encoding a reverse transcriptase-related protein.";  
RL Nature 321:625-628(1986).  
CC -!- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF  
CC PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS  
CC LABORATORIES, BELONGING TO THE LINE-1 FAMILY.  
DR PIR: A25313; GNHUI1.  
DR InterPro: IPR005135; Exo\_endo\_phos.  
DR InterPro: IPR000300; IPPC.  
DR InterPro: IPR000477; RVTse.  
DR Pfam: PF000078; rvt; 1.  
DR Pfam: PF03372; Exo\_endo\_phos; 1.  
DR SMART: SM00128; IPPC; 1.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 1259 AA; 147112 MW; F3BC18A3803919F5 CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 1259  
Score: 5139.00 Matches: 1050  
Percent Similarity: 79.00% Conservative: 71  
Best Local Similarity: 74.00% Mismatches: 132  
Query Match: 9.82% Indels: 173  
DB: 1 Gaps: 7

US-10-083-853B-2 (1-29921) x LN1\_HUMAN (1-1259)

QY	3434	GGATCAAAATTACACATAATAATATTAACTTTAAATGTAAATGGGCTAAATTCGCCCAATT	3493
Db	2	GlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnAlaProIle	21
QY	3494	AAAAGCACAGACTGCRAATTGGATAAAGAGTCAAGACCCTCAGTGCTGTATTTCAG	3553
Db	22	LysArgHisArgLeuAlaAsnTrpIleLysSerGlnAspProSerValCysCysIleGln	41
QY	3554	GAGGCCCATCTCATATGAAGAAGACACACATAGCCCTCAAAYTAAGGGGATGGAGGAAGTT	3613
Db	42	GluThrHisLeuThrCysArgaspThrHisArgLeuLysIleLysGlyTrpArgAsnIle	61
QY	3614	TACAAGTAATAAGAAAAACAAAAAAGCAGGGTTGCAATCTCTAGTCTCGATAAAA	3673
Db	62	TyrGlnAlaAsnGlyLysGln--LysLysAlaGlyValAlaIleLeuValSerAspLys	80
QY	3674	ACAGACTTTAAACCAACAAGATCAAAAGAGACAAAGAAGGCCATTCATATCGTTAAAG	3733
Db	81	ThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyRileMetValLys	100
QY	3734	GCATCAATGGAACAAGAAGACTAACTATCTCTAAATATATACATGCACCCAATFACAGGAGCA	3793
Db	101	GlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyRAlaProAsnThrGlyAla	120
QY	3794	CCCAGATTCAATAAGCAAGTCTTAGAGACCTTACAAGAGACTTTTGACTCCCCACACAATA	3853
Db	121	ProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHisThrIle	140
QY	3854	ATAGTGGGAGTCTAAATAATAATAGACACTTTAACACCCCCACTGCCAATATTAGGCAGA	3913
Db	141	IleMetGly-----AspPheAsnThrProLeuSerThrLeuAspArg	154
QY	3914	TCAATGAGCAGACAGAAAATTAACAAGGATATACAGGAGTTGAACTGAGCTCGGACCAAGCG	3973
Db	155	SerThrArgGlnLysIleAsnLysAspIleGlnGluLeuAsnSerAlaLeuHisGlnAla	174
QY	3974	GACCTAATAGATATCTACAGAACTCCCACCCCCAAATCAACAGAATATACACTCTTCTCA	4033
Db	175	AspLeuIleAspIleTyRArgThrLeuHisProLysSerThrGluTyThrPhePheSer	194
QY	4034	GCATCATTACACCTATTTTAAATTCACCATTAATTTTAACTAAACACTCCCTCAGC	4093
Db	195	AlaProHisThrTyRSerLysThrAspHisIleLeuGlySerLysThrLeuLeuSer	214
QY	4094	AAATGCAAAAGAACAGAAATCTTAACAACAGCTCTCTCAGACTACAGTGAATCTATTTA	4153
Db	215	LysCysLysArgThrGluIleIleThrAsnCysLeuSerAspHisSerAlaIleLysLeu	234
QY	4154	GAACTCAGAATTAAGAAACTACTCAAAATCACACAACCTACATGAGAACTGAACAACCTG	4213
Db	235	GluLeuArgIleLysLysLeuThrGlnAsnHisSerThrThrTrpLysLeuAsnAsnLeu	254
QY	4214	CTCCTGAATGACTACTGGGTAAATTAACAATAATGAAGCAAAAATAAAGATGTTCTTTGAA	4273
Db	255	LeuLeuAsnAspTyRTrpValHisAsnGluMetLysAlaGluIleLysLysPhePheGlu	274
QY	4274	ACCAATGAGAACAAAGACACAATGPACAGAAATCTCTGGGGCATATTTAAAGCAGTGTGT	4333
Db	275	ThrAsnGluAsnLysAspThrThrTyRGlnAsnLeuTrpAspThrAlaLysAlaValCys	294
QY	4334	AGAGGGAATTTTATAGCTACTAGTCCCTACAAGAGAAGCAGGAATATCTTAAANTAGAC	4393
Db	295	ArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIleAsp	314
QY	4394	ACCTTAAACATCAATTAAGAAGACTAGAGAAGAAAGCAACAATTTCAAAGCTAGC	4453
Db	315	ThrLeuIleSerGlnLeuLysGlnLeuGluLysGlnGlnGlnThrAsnSerLysAlaSer	334
QY	4454	AGAAGACAAGAAATAACTAAGATCAGACGAACCTGAAGGAGATAGAGACACAAAAAGCC	4513
Db	335	ArgArgGlnGluIleIleLysIleArgAlaGluLeuLysGluIleGluThrGlnLysThr	354
QY	4514	CTTCAAAATAAATCAATGAATCCAGGAGCTGGTTTTTTTCAAAAAGATCAGCAAAAT---	4569

Db	564	lyPheIleProAlaMetGlnGlyTrpPheAsnIleArgLysSerIleAsnIleIleGln	584	QY	6718	AGAAAGCTGAACCTGGATCCCTTCCTTACACCTTATACAAAGCTTAACCTCAAGATGAATT	6777
QY	5647	ACATAAACAACAACCAATGACAAAAACACATGATTATCTCAATAGATGACAAAAAGCCT	5706	Db	944	ArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAsnSerArgTrpIle	963
Db	584	isIleAsnArgThrLysAspThrAsnHisMetIleIleSerIleAspAlaGluLysAlaP	604	QY	6778	AAAGACTTAATAAGACATATAAACCAATAAAACCCCA-GAAGAAACCTTAGCAATACC	6836
QY	5707	TTGTCAAAATPACACGCCCTTCATGCTCAAAAAATCTCAGTAAATAGGTATGATGGAA	5766	Db	964	LysAspLeuAsnValArgProLysThrIleLysThrLeuGluLysAsnLeuGlyAsnThr	983
Db	604	heAspLysIleGlnGlnProPheMetLeuLysProLeuAsnLysLeuGlyIleAspGlyT	624	QY	6837	ATTCAGATATGCACATGGCAAGACTTCATCACTAAACACCAAAACCAATGGCAACA	6896
QY	5767	TGTATCTCAAAATAAAGAGCTATTATTATAC-AAACCCACAGCAATATCATCTAATG	5825	Db	984	IleGlnAspIleGlyMetGlyLysAspPheMetThrLysThrProLysAlaMetAlaThr	1003
Db	624	hrTyrrLeuLysIleIleArgAlaIleTyrrAspLysProThrAlaAsnIleIleLeuAsnG	644	QY	6897	AAAGCAAAATAGACAAGTGGGATCTGATTAAATATAGAGCTCTGCACAGCAAAAAA	6956
QY	5826	GGCAAAACTGGAAGCATTCCTTTGAGAACTGGCACAGACAGAGTGGCTCTCTCAC	5885	Db	1004	LysValLysIleAspArgTrpAspLeuIleLysLeuLysSerPheCysThrAla-LysG1	1023
Db	644	lyGlnLysLeuGluAlaProProLeuLysThrlyThrArgGlnGlyCysProLeuSerP	664	QY	6957	AACTGTCATCAGAGTGAACAGCAACCTACAGAAATGGGAGAAATTTTTCGAATCTATCG	7016
QY	5886	CACCTCTATTCAAGATCTATTGGAAGTTCTTGCCAGGGCAATPACGGCAATAGAAAGAA	5945	Db	1023	uThrThrIleArgValAsnArgGlnProThrLysTrpGluLysIlePheAlaIleTyrrSe	1043
Db	664	roLeuLeuProAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLysGluI	684	QY	7017	ATCTGACAAAGCGCTAATATCCAGAGATCTACGAAGAACTTAAACAAATTTACAGAAAA	7076
QY	5946	TAAAGGTATTCAATAGAAAGAGAGAGACTCATATTGCTCTGTTTGCAGATGACATGT	6005	Db	1043	rSerAspLysGlyLeuIleSerArgIleTyrrAsnGluLeuLysGlnIleTyrrLysLysL	1063
Db	684	leLysGlyIleGlnLeuGlyLysGluValLysLeuSerLeuPheAlaAspMetI	704	QY	7077	---AAACAACCCCGCTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAAAGAAGA	7132
QY	6006	TTGTATATTAGAAAACCCCATCTCTCAGGGCCCAAAACCTCTTAAGCTGATAGAACA	6065	Db	1063	sThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisProSerLysGluAs	1083
Db	704	leValTyrrLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleSerAsnP	724	QY	7133	CATTATGAGCCCAACAAACATATGAAAAACCTCATCATCATCTGGTGGTGTAGAGAAAT	7192
QY	6066	TCAGCAAAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAAGCATTTCTTACGCCCA	6125	Db	1083	pIleTyrrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleArgGluMe	1103
Db	724	heSerLysValSerGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeuTyrrThra	744	QY	7193	GCAAAACAAACCCACAGTGCATACCATCTCATGCTAGTGTAGATGGTGTATCACTAAAA	7252
QY	6126	ATAATAGACAAACAGAGACCAATCATGAGTGAACCTCATTCACAATGCTACAAGA	6185	Db	1103	tGlnIleLysThrThrMetArgTyrrHisLeuThrProValArgMetAlaIleIleLysL	1123
Db	744	snAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIleAlaSerLys	764	QY	7253	GTGAGAAACAAACAAATGCTGGAGAGATGTGGAGAAATAGGAACACTTTTCCACTGTTG	7312
QY	6186	GAATAAAATACCTAGGAATACAACTTACAAGGACACAGTAGGAACCTCTCAAGGAGA	6245	Db	1123	sSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTr	1143
Db	764	rgIleLysTyrrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysGluAsnT	784	QY	7313	GTGGGAATGTAATATTAGTCAACCATTTGGGAAGACAGTGTGGAGATTCCTTAAAGATCT	7372
QY	6246	ACAAACCATGTATCAAGGAATAAAGAGAGACACAAACAAATGGAAAAACATTCATGCT	6305	Db	1143	pTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLe	1163
Db	784	yrLysProLeuLeuAsnGluIleLysGluAspThrAsnLysTrpLysAsnIleProCys	804	QY	7373	AGAACCCAGAAATATCATTTGACCCCACTCCCATCTACGATATATATCCCAAGGAATA	7432
QY	6306	CACAGATAGTAAAGATCAT-----GAAATGCCATCTGCCCCCAAGTAATATAGAT	6358	Db	1163	uGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrrProLysAspTy	1183
Db	804	erTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleTyrrArgp	824	QY	7433	TAAATCATTTCTATTATAAGACACATGTCACACATATGTTTATTTGCAGCACTGTATCAAT	7492
QY	6359	TCAGTGTACCCCATCAAGCTACCATGACTTCTTTCAGAGAAATTTGGAAAAACAACTT	6418	Db	1183	rLysSerCysCysTyrrLysAspThrCysThrArgMetPheIleAlaLeuPheThrI1	1203
Db	824	heAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysThrL	844	QY	7493	AGCAAGACTTGGAAACCAACCCCAATGTCCATCAGTGTAGATGTGATGATAAGAAACATG	7552
QY	6419	TAAATTTTCAATGGAACCAAAAAAGAGCCCAAGCCCAAGACAAATCTTAAGCAAAAG	6478	Db	1203	eAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLysMetTr	1223
Db	844	eulysPheIleTrpAsnGln-LysArgAlaHisIleAlaLysSerThrLeuSerGlnLys	863	QY	7553	GCATATATACACCATGAAATCTATCTGCAGCCATATAAAG-GATGAGTTCATGCTTTCG	7611
QY	6479	AACAAAGCTGGAGTATCATGCTACCTGACTTAAACTATATATAGGCTACAGTAACC	6538	Db	1223	pHisIleTyrrThrMetGluTyrrAlaAlaIleLysAsnAspGluProMetSerProVa	1243
Db	864	AsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrrTyrrLysAlaThrValThr	883	QY	7612	AGAGATATGATGAAGCTGGAACCATCATCTTCTCAGCAAACTAACCAAA 7660	
QY	6539	AAACCTGCATGGTGGTACCAAAAACAGATATATAGACAAATGGAAACAGACAGAGACC	6598	Db	1243	lGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGln 1259	
Db	884	LysThrAlaTrpTyrrTyrrGlnAsnArgAspIleAspGlnTrpAsnArgThrGluPro	903	RESULT 2			
QY	6599	TCAGAAATTAC-ACTGCAATCTACATCCATCTGATCTTTGACAAACCTGACAAAAACAAG	6657	LINE1_NYCCO	STANDARD;	PRT;	1260 AA.
Db	904	SerGluIleMetProHisIleTyrrAsnTyrrLeuIlePheAspLysProGluLysAsnLys	923	AC	08548;		
QY	6658	CAATGGAAAAGGATCCCTATTATTAATGCTGTGGAAAAACCTGGCTAGCCATATGC	6717	DT	01-AUG-1988 (Rel. 08, Created)		
Db	924	GlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaIleCys	943	DT	01-AUG-1988 (Rel. 08, Last sequence update)		
				DE	01-AUG-1988 (Rel. 08, Last annotation update)		
					LINE-1 reverse transcriptase homolog.		

OS Nycticebus coucang (Slow loris).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Strepsirhini; Loridae; Nycticebus.  
 OX NCBI\_TaxID=9470;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=86230917; PubMed=2423883;  
 RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;  
 RT "LI family of repetitive DNA sequences in primates may be derived  
 from a sequence encoding a reverse transcriptase-related protein.";  
 RL Nature 321:625-628(1986).  
 CC -1- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF  
 SIX SEQUENCES, DETERMINED BY THESE AUTHORS BUT NOT SHOWN,  
 BELONGING TO THE LINE-1 FAMILY.  
 CC PIR; B25313; GNRL1.  
 DR HSSP; P27695; 1HD7.  
 DR InterPro; IPR005135; Exo\_endo\_phos.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00078; rvt; 1.  
 DR Pfam; PF03372; Exo\_endo\_phos; 1.  
 DR RNA-directed DNA polymerase.  
 KW RNA-directed DNA polymerase.  
 SQ SEQUENCE 1260 AA; 147042 MW; 7A6803DE471F7253 CRC64;

Alignment Scores:  
 Pred. No.: 6,62e-256 Length: 1260  
 Score: 3658.50 Matches: 743  
 Percent Similarity: 66.90% Conservative: 207  
 Best Local Similarity: 52.32% Mismatches: 303  
 Query Match: 6.99% Indels: 174  
 DB: 1 Gaps: 9

US-10-083-853b-2 (1-29921) x LINI\_NYCCO (1-1260)

QY 3434 GGATCAAAATTCACACATATAATTAACCTTAAATGGCTAAATCCCAATT 3493  
 DB 2 GlyLeuSerGlyLeuSerIlePheSerIleAsnValAsnGlyLeuAsnCysProLeu 21  
 QY 3494 AAAAGACACAGACTGGCAAAATGGATAAAGAGCTCAAGACCCCATGCTGTGTTTCCAG 3553  
 DB 22 LysArgHisArgLeuAlaAspTrpIleGlnLysLeuLysProAspIleCysCysIleGln 41  
 QY 3554 GAGGCCCTCTCACATGAAGACACACATAGGCTCAAAATAAAGGATGGAGGAATT 3613  
 DB 42 GluSerHisLeuThrLeuLysAspLysTyrArgLeuLysValLysGlyTrpSerSerIle 61  
 QY 3614 TACCAAGTAATGGAAACAAAAAAGAGGGGTGGCAATCTGTCTCTGATAAA 3673  
 DB 62 PheGlnAlaAsnGlyLysGln---LysLysAlaGlyIleAlaIleLeuPheAlaAspAla 80  
 QY 3674 ACAGACTTTAAACCAACAAAGATCAAAAGACAGACAAAGGCCATTACATATGTTAAAG 3733  
 DB 81 IleGlyPheLysProThrLysIleArgLysAspLysAspGlyHisPheIlePheValLys 100  
 QY 3734 GCATCAATGGACACAGAGAGCTACTACTCTTAATATACATGCACCCCAATACAGGACA 3793  
 DB 101 GlyAsnThrGlnTyrAspLysIleSerIleIleAsnIleTyrAlaProAsnHisAsnAla 120  
 QY 3794 CCCAGATTCAAAAGCAAGTCTTTAGAGACCTTACAAAGAGACTTTTGACTCCCAACAATA 3853  
 DB 121 ProGlnPheIleArgGluThrLeuThrAspMetSerAsnLeuIleSerThrSerIle 140  
 QY 3854 ATAGTGGAGTCTAAATAATAATAGACACTTTAAACCCCACTTAAACCCCACTTAAAGGACA 3913  
 DB 141 ValValGly-----AspPheAsnThrProLeuAlaValLeuAspArg 154  
 QY 3914 TCAATGACACAGAAATTAACAAGGATATCCAGGAGTTGAAGTCTGCTGGACCAACCG 3973  
 DB 155 SerSerLysLysLysLeuSerLysGluIleLeuAspLeuAsnSerThrIleGlnHisLeu 174  
 QY 3974 GACCTAAATAGATATCTACAGAACTCCCAACCCCAATCAACAGATATACACTCTTCTCA 4033  
 DB 175 AspLeuThrAspIleTyrArgThrPheHisProAsnLysThrGluTyrThrPhePheSer 194

QY 4034 GCATCACATTACACCTATTTTAAATTTGACCATGTAAATTTTAAAGTAAACACATCTCTCAGC 4093  
 DB 195 SerAlaHisGlyThrTyrSerLysIleAspHisIleLeuGlyHisLysSerAsnLeuSer 214  
 QY 4094 AAATCCAAAGAACAGAAATCTTAACAACACAGCTCTCAGACTACAGTCAATCTATTTA 4153  
 DB 215 LysPheLysLysIleGluIleProCysIlePheSerAspHisHisGlyIleLysVal 234  
 QY 4154 GAACTCAGAATTAACAAATCACTCAAAATCAACAACATACATGGAATCAACACCTG 4213  
 DB 235 GluLeuAsnAsnAsnArgAsnLeuHisThrHisThrLysThrTrpLysLeuAsnAsnLeu 254  
 QY 4214 CTCCTGAATGACTACTGGTAAATAACAATAAGAGCAAAATAAAGATGCTCTTGA 4273  
 DB 255 MetLeuLysAspThrTrpValIleAspGluIleLysLysGluIleThrLysPheLeuGlu 274  
 QY 4274 ACCAATGAGAACAAAGACACAAATGTACCAGATCTCTGGGCATATTTTAAACAGCTGT 4333  
 DB 275 GlnAsnAsnAsnGlnAspThrAsnTyrGlnAsnLeuTrpAspThrAlaLysAlaValLeu 294  
 QY 4334 AGAGGGAATTTATAGCACTAGATCGCTACAGAGAAAGACGAGCAAAATATCTAAAATAGAC 4393  
 DB 295 ArgGlyLysPheIleAlaLeuGlnAlaPheLeuLysLysThrGluArgGluGluValAsn 314  
 QY 4394 ACCTTAACATCACAAATTAAGAAGACTAGAGAGAAAGACAAACAAATTCAAAAGCTAGC 4453  
 DB 315 AsnLeuMetGlyHisLeuLysGlnLeuGluLysGluGluHisSerAsnProLysProser 334  
 QY 4454 AGAGACAAAGAAATTAAGATCAGACAGCACTGAAGGAGATAGACACACAAAGGCC 4513  
 DB 335 ArgArgLysGluIleThrLysIleArgAlaGluLeuAsnGluIleGluAsnLysArgIle 354  
 QY 4514 CTTCAATTAATCAATGAATCCAGGAGCTGGTTTTTGAAGAAGATCAGCAAAATAGA --- 4570  
 DB 355 IleGlnGln-IleAsnLysSerLysSerTrpPheGluLysIleAsnLysIleAspLys 374  
 QY 4571 -CCACTAGACAGACTTAATAAG 4629  
 DB 374 sProLeuAlaAsnLeuThrArgLysLysArgValLysSerLeuIleSerSerIleArgAs 394  
 QY 4630 TGATAAAGGGATATCACCCCGGATCCCAAGAGATCAAAACTATTATCAGAGATATTTA 4689  
 DB 394 nGlyAsnAspGluIleThrThrAspProSerGluIleGlnLysIleLeuAsnGluTyrTy 414  
 QY 4690 TAAACACCTCTATGCAATAAATAGAAAATCTAGAAAATGGATAAATCTCTCGACAC 4749  
 DB 414 rLysLysLeuTyrSerHisLysTyrGluAsnLeuLysGluIleAspGlnTyrLeuGlu-- 433  
 QY 4750 ATATGTAGCCTGTATGGACCTTGCGGGGACAGAAAGGGGGTGAATGCAGAAATAAAA 4809  
 DB 433 ----- 433  
 QY 4810 GACAAAGACAAAGAGTATGTTGGAGTAGGGGTGAGGGGCACTTGCCTCTAATGGA 4869  
 DB 433 ----- 433  
 QY 4870 CAAGGGCCCTGAGCTTTACACCACCTCTGATTTATTAGGCAAAAGAGATAGCAGAGG 4929  
 DB 433 ----- 433  
 QY 4930 GTGAGTTGGAAGAGAGGTGAGCTGTTAGGTCAGAGTAGGCTGCAAGACTGCATTCCT 4989  
 DB 434 -----AlacysHis----- 436  
 QY 4990 CAAACAATAGGCTCTAGATGTCCAGTAGATAACCTCAAGGAGCCAGTGCAGGGAGTGA 5049  
 DB 436 ----- 436  
 QY 5050 TGGCCCTCAGCAAACTTCTAGGGCAGGCACAGAAGTAAGTTTGCCACATCTCTGATTC 5109  
 DB 436 ----- 436  
 QY 5110 ACGATAAACAGTTTGTCTGTTTGTGATCAAGTAGCCTCCAGTGAATGCTGAGTTGGTCATGA 5169



```
QY 7313 GTGGGAATGTAATAGTTCACCAATTGTGGAAGACAGATGTGGAGATTCTTAAAGGATCT 7372
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1143 tTpGluCysProLeuIleArgSerPheTrpLysAspValTrpArgIleLeuArgAspLe 1163

QY 7373 AGAACCAATATCATTGTGACCAGCAATCCCAATCTACTAGTATATACCCAAAGGAATA 7432
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1163 uLysIleAspProPheAspProIleIleProLeuLeuGlyLeuTyProGluAspGl 1183

QY 7433 TAAATCATCTTATTATAAGACACATGCACACATATGTTTATTCAGCAGTCATCAACAT 7492
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1183 nLysSerGlnTrpAsnLysAspIleCysThrArgMetPheIleAlaGlnPheIleIl 1203

QY 7493 AGCAAGACTTGGAAACCAACCAATGTCCATCATGATAGACTGGGATAAGAAAACATG 7552
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1203 eAlaLysSerTrpLysLysProLysCysProSerThrHisGluTrpThrSerLysLeuTr 1223

QY 7553 GCACATATACACCATGAATACTATGACGACCAATAAAGGAT-----GAGTTCATGTCCTT 7608
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1223 pTyMetTyThrMetGluTyTyAlaAlaLeuLysLysAspGlyAspPheThrSerPh 1243

QY 7609 TGCAGAGATATGGATGAAGCTGGAACCATCATCTCTCAGCAAACTAACACAA 7660
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1243 eMetPheThrTrpMetGluLeuGluHisIleLeuLeuSerLysValSerGln 1260

RESULT 3
POL2_MOUSE
ID POL2_MOUSE STANDARD; PRT; 1300 AA.
AC P11369;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retrovirus-related POL polyprotein [Contains: Reverse transcriptase
DE (EC 2.7.7.49); Endonuclease].
GN POL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064284; PubMed=3023821;
RA Leob D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,
RA Edgell M.H., Hutchison C.A. III;
RT "The sequence of a large L1MD element reveals a tandemly repeated 5'
RT end and several features found in retrotransposons.";
RL Mol. Cell. Biol. 6:168-182(1986).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + [DNA](N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13002; AAA66024.1; ALT_INIT.
DR PIR: B24906; GNMSLL.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF03372; Exo_endo_phos; 1.
KW Hydrolase; Transferase; RNA-directed DNA polymerase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 1300 AA; 151829 MW; 9EB8108493EB9635 CRC64;

Alignment Scores:
Pred. No.: 5.88e-248 Length: 1300
Score: 3548.50 Matches: 729
Percent Similarity: 65.39% Conservative: 208
Best Local Similarity: 50.87% Mismatches: 329
```









```
Db 974 LysThrLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe 993
Qy 29439 ATGTCTAAACACCAAAAGCAATGGCAACAAAGCCAAAATTCACAAATGGATCTAATT 29380
RL :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 994 MetThrLysThrProLysAlaMetAlaThrLysValLysIleAspArgTTrpAspLeuIle 1013
Qy 29379 AACTAAAGAGCTTCTCCACAGCAAAAGAAACACATCACATGAGAGTGAACAGGCCAACTTACA 29320
Db 1014 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 1033
Qy 29319 GAATGGAGAGAAATTTTGGCAACCTACTCATTTGACAAAGGGCTAATATCCAGATCTAC 29260
Db 1034 LysTrpGluLysIlePheAlaIleTyrSerSerAspLysGlyLeuIleSerArgIleTyr 1053
Qy 29259 AATGAATCAACAAATTTTACAGAAAAAAGAAAGAACCCCATCAAAAGAGTGGTGAAG 29200
Db 1054 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLys 1073
Qy 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGACGCCCAAAACACATCAAAAAA 29140
Db 1074 AspMetAsnArgHisProSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLys 1093
Qy 29139 TACTCATCATCTGGCCATCAGAGAAATGCAATCAAAACCAACCAATGAGATCCATCTC 29080
Db 1094 CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 1113
Qy 29079 ACACCACTTAGAATGGCAATCATTAAAGTCAAGAAACACAGGCTCTGAGAGATGT 29020
Db 1114 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 1133
Qy 29019 GGAGAAATAGAACACTTTTACACTGTGTGGTGGACTGTGTAAGTGTCAACCATTTGTGG 28960
Db 1134 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 1153
Qy 28959 AAGTCAGTGTGGTATCTCTCAGGATCTAGAACTAGAAATACATTTGACCCAGCATC 28900
Db 1154 LysSerValTrpArgPheLeuArgAspLeuGluLeuIleProPheAspProAlaIle 1173
Qy 28899 CCATTACTGGTATATACCAAGAGGATTATAATCATGCTGCTATTAAGACACACAGCAC 28840
Db 1174 ProLeuLeuGlyIleTyrProLysAspTyrLysSerCysTyrLysAspThrCysThr 1193
Qy 28839 TGTATGTTTATAGCAGCATTTCACAAATAGCAAGACTTGGAAACCAACCTAAATGTCCA 28780
Db 1194 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1213
Qy 28779 ACAACGATAGACTGGATTAGAAAAATGTGGCAGATATACACCATGGATGGAATCTAGCAGCC 28720
Db 1214 ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 1233
Qy 28719 ATAAAAATGATGTCGTCTCTTTGTAGGGACATGGATGAAGCTGGAACCATCATTT 28660
Db 1234 IleLysAsnAspGluProMetSerProValGlyThrTrpMetLysLeuGluThrIleIle 1253
Qy 28659 CTCAGCAACTATACAA 28642
Db 1254 LeuSerLysLeuSerGln 1259
```

## RESULT 5

```
LIN1_NYCCO STANDARD; PRT; 1260 AA.
AC P08548;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE LINE-1 reverse transcriptase homolog.
OS Nycticebus coucang (Slow loris).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Loridae; Nycticebus.
OX NCBI_TaxID=9470;
RN [1]
RP SEQUENCE.
RX MEDLINE=86230917; PubMed=2423883;
```

```
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "L1 family of repetitive DNA sequences in primates may be derived
RL Nature 321:625-628(1986).
CC -1- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC SIX SEQUENCES, DETERMINED BY THESE AUTHORS BUT NOT SHOWN,
CC BELONGING TO THE LINE-1 FAMILY.
DR PIR; B25313; GNURL1.
DR HSP; P27695; 1HD7.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1260 AA; 147042 MW; 7A6803DF471F7253 CRC64;
```

## Alignment Scores:

```
Pred. No.: 1.56e-98 Length: 1260
Score: 1479.50 Matches: 265
Percent Similarity: 78.22% Conservative: 69
Best Local Similarity: 62.06% Mismatches: 92
Query Match: 2.82% Indels: 1
DB: 1 Gaps: 1
```

US-10-083-853B-2 (1-29921) x LIN1\_NYCCO (1-1260)

```
Qy 29919 ACTTCTCTCAGCAATTTGGAAGAACTACTTTAAAGTTCATGTGGAACCAAAAGAGCC 29860
Db 834 SerTyrPheLysAspLeuGluLysIleLeuHisPheIleTrpAsnGlnLysLysPro 853
Qy 29859 CACATCCCAAGTCAATCTCTAAAGCCAAAGAAAGCAAGCTGAGGCATCATCTACCTGCAC 29800
Db 854 GlnIleAlaLysThrLeuLeuSerAsnLysAlaGlyGlyIleThrLeuProAsp 873
Qy 29799 TTCAAACTACATCAACAGGTACAGTAACCAAAACACATGCTGCTGTCACCAAAACAGA 29740
Db 874 LeuArgLeuTyrTyrLysSerIleValIleLysThrAlaTrpTyrTrpHisLysAsnArg 893
Qy 29739 GATATCAACCAATGGAACAGACAGAGCCCTCGAATAATATCGGCATATCTACCACTAT 29680
Db 894 GluValAspValTrpAsnArgIleGluAsnGlnGluMetAspProAlaThrTyrHisTyr 913
Qy 29679 CTGATCTTTGACAAACCTGAGAAACCAAGCAAGTGGGAAGAGTTCCTCTATTATATAA 29620
Db 914 LeuIlePheAspLysProIleLysAsnIleGlnTrpGlyLysAspSerLeuPheAsnLys 933
Qy 29619 TGGTGTGAGAAAAGTGGCTAGCCCTATGTAGAAAGCTGAAAGTGGATCCCTTCTTACA 29560
Db 934 TrpCysTrpValAsnTrpLeuAlaIleCysArgArgLeuLysLeuAspProHisLeuSer 953
Qy 29559 CCTTATACAAAATTAATTTCAAGATGGATTAAAGACTTAATGTTAGACCTAAACCATTA 29500
Db 954 ProLeuThrLysIleAspSerHisTrpIleLysAspLeuAsnLeuArgHisGluThrIle 973
Qy 29499 AAATCCCTAGAGAAAACCCAGCAATACCATTACGACATGAGCATGGCAGGAGGATTC 29440
Db 974 LysIleLeuGluGluSerAlaGlyLysThrLeuGluGlyIleSerLeuGlyGluTyrPhe 993
Qy 29439 ATGTCTAAACACCAAAAGCAATGGGCAACAAACCAAAATTCACAATGGGATCTAATT 29380
Db 994 MetArgArgThrProGlnAlaIleGluAlaValSerLysIleHisTyrTrpAspLeuIle 1013
Qy 29379 AAATAAGAGCTTCTGCACAGCAAAAGAAACTACCATCAGAGTGAACAGCACTTACA 29320
Db 1014 LysLeuLysSerPheCysThrAlaLysAsnIleValSerLysAlaSerArgGlnProSer 1033
Qy 29319 GAATGGAGAGAAATTTTGGCAACCTACTCTTTTGACAAAGGGCTAATATCCAGATCTAC 29260
Db 1034 GluTrpGluLysIlePheAlaGlyTyrThrSerAspLysGlyLeuIleThrArgIleHis 1053
Qy 29259 AATGAATCAACAAATTTTACAGAAAAAAGAAACCCCATCAAAAGAGTGGTGAAG 29200
Db 1054 ArgGluLeuLysHisIleAsnLysLysArgThrArgAspProIleSerGlyTyrAlaArg 1073
```















CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U14572; ; NOT\_ANNOTATED\_CDS.

KW Hypothetical protein.

FT DOMAIN 1 97 FRAME-1.  
 FT DOMAIN 101 196 FRAME-2.  
 FT DOMAIN 200 295 FRAME-3.  
 FT DOMAIN 299 395 FRAME-4.  
 FT DOMAIN 399 494 FRAME-5.  
 FT DOMAIN 498 593 FRAME-6.  
 SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:  
 Pred. No.: 4,35e-20 Length: 593  
 Score: 392.00 Matches: 77  
 Percent Similarity: 82.29% Conservative: 2  
 Best Local Similarity: 80.21% Mismatches: 17  
 Query Match: 0.75% Indels: 1  
 DB: 1 Gaps: 0

US-10-083-853B-2 (1-29921) x ALU6\_HUMAN (1-593)

QY 10343 TTTTGTGAGATGGAGTTTCAGTCTTCCAGGCTAGAGTCAATGGTGATCCC 10402  
 DB 299 PhePheLeuArgSerPheAlaLeuValAlaGlnAlaGlnTrpArgAspLeu 318  
 QY 10403 GGGTCACTGCAACCTCCACCTCCCGGGTCAAGTGTCTCTCCCTCAGCTCCCAAGT 10462  
 DB 319 GlySerProGlnProProGlyPheLysArgPheSerCysLeuSerLeuProSer 338  
 QY 10463 AGCTGGGACTACAGCAGCTGCCACCGCTGCTTAATTTTGTATTTTAGTAGAGAT 10522  
 DB 339 SerTrpAspTyrArgHisAlaProProArgProAlaAsnPheCysIlePheSerArgAsp 358  
 QY 10523 GGGTT-TCCCCACGTTGCCAGGCTGGTCTTGAACCTCTGACCTCAGCTGATCTGCCCTGC 10581  
 DB 359 GlyValSerProCysTrpSerGlyTrpSerArgTrpProAspLeuArg\*\*\*SerAlaArg 378  
 QY 10582 CTCAGCCPCCCAAAATGCTGGGATTACAGGCATGAGCCACCTGCCGCCG 10629  
 DB 379 LeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgPro 394

RESULT 11  
 Y011\_MOUSE  
 ID Y011\_MOUSE STANDARD; PRT; 379 AA.  
 AC P11260;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-JUL-1989 (Rel. 11, Last annotation update)  
 DE Hypothetical protein ORF-1137.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINF=87064284; PubMed=3023821;  
 RA Loeb D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,  
 RA Edgell M.H., Hutchison C.A. III;  
 FT "The sequence of a large L1MD element reveals a tandemly repeated 5'  
 FT end and several features found in retrotransposons."  
 RL Mol. Cell. Biol. 6:168-182(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M13002; AAA66023.1; ALT\_INIT.

DR PIR; A24906; QOMSL.

DR InterPro; IPR004244; Transposase\_22.

DR Pfam; PF02994; transposase\_22; 1.

KW Hypothetical protein.

SQ SEQUENCE 379 AA; 43788 MW; F4DDE312807644D3 CRC64;

Alignment Scores:  
 Pred. No.: 8,74e-20 Length: 379  
 Score: 387.00 Matches: 115  
 Percent Similarity: 46.74% Conservative: 64  
 Best Local Similarity: 30.03% Mismatches: 139  
 Query Match: 0.74% Indels: 65  
 DB: 1 Gaps: 9

US-10-083-853B-2 (1-29921) x Y011\_MOUSE (1-379)

QY 2301 TCAACACATCAAAAGACC-----AAAGTAAATATAAAC 2333  
 DB 10 ThrValThrThrLysThrAlaSerPheArgAspTyrGlnMetAlaLysGlyLys----- 27  
 QY 2334 CAAAAGATGGGAAAAACACAGTGCAGAACACTGAAATTCACAAACACCAAGACTCTCT 2393  
 DB 28 -----ArgLysAsnProThrAsnArgAsnGlnAsp 37  
 QY 2394 TCTCAACCAAGGATCACAACTCTCCGCCACA-----AGGACACAAACACAG 2441  
 DB 38 HisSerProSerSerGluArgSerThrProThrProProSerProGlyHisProAsnThr 57  
 QY 2442 ATGGAGATGAGTTTGAGGAATTCACAGAAGTAGGCTTCAGAAAGTGGTGAATACACAA 2501  
 DB 58 ThrGluAsnLeuAspProAspLeuLysThr-----PheLeuMetMetMetIleGluAsp 75  
 QY 2502 -----TCCTCCGAGCTAAAGAGAGATGTTCTAACCCATGCAAGGAAGCT 2546  
 DB 76 IleLysLysAspPheHisLysSerLysSerLysLeuGlnGluSerThrAlaLysGluLeu 95  
 QY 2547 AAGAACCTTGAAGAAAGCTTAGATCAATTGCTAACTAGATAATACAGTGTAGAGAGAAG 2606  
 DB 96 GlnAlaLeuLysGluLysGlnGluAsnThrAlaLysGlnValMetGluMetAsnLysThr 115  
 QY 2607 ATAAATGACCTG-----ATG 2621  
 DB 116 IleLeuGluLeuLysGlyGluValAspThrIleLysLysThrGlnSerGluAlaThrLeu 135  
 QY 2622 GAGCTGAAAAACGCAAGACAAAGACTTCATGAAGCATACACAGCTTCAATAGCCAAATC 2681  
 DB 136 GluIleGluThrLeuGlyLysArgSerGlyThrIleAspAlaSerIleSerAsnArgIle 155  
 QY 2682 GATCAACGACGAAAGAGATATCATGATTGAAGATCAATTAATAAAGAAAGAGTGAGAA 2741  
 DB 156 GlnGluMetGluGluArgIleSerGlyAlaGluAspSerIleGluAsn----- 171  
 QY 2742 CACAAGATTACAGAAAAAGAGTGAAGAAAGAAC-----AAACAA 2780  
 DB 172 -----IleAspThrThrValLysGluAsnThrLysCysLysArgIleLeuThrGln 188  
 QY 2781 AGCCTCCAGAAATATTGGGACTATGTGAAAGACCAAAATCTACATTTGATGGTCCCC 2840  
 DB 189 AsnIleGlnValIleGlnAspThrMetArgArgProAsnLeuArgIleIleGlyLeuAsp 208  
 QY 2841 CAAGTGTATGGGAGATGGATCACTGAGTGGAAACACTCTTCAGGGGTATTATCCAGGAG 2900  
 DB 209 GluAsnGluAspPheGlnLeuLysGlyProAlaAsnIlePheAsnLysIleIleGluGlu 228

QY 2901 AATTCCCC---ATCTATCAGGGCAGGCCAACATTCAGAAATATGGAGAACAC 2957  
 Db 229 AsnPhenileLysGluMetProMetile--IleGlnGluAlaTrpArgThrP 248  
 QY 2958 CATAAAGTACTCTCGAGAGACATCCCAACACATATCTTCAGATTCACCAAGG 3017  
 Db 248 roAsnArgLeuAspGlnLysArgAsnSerArgHisIleIleIleArgThrAsnA 268  
 QY 3018 TTGAATAGAGAAAAATGTTAAGGCAGCAGAGAGAAAGTGGTTCACCAAAAG 3077  
 Db 268 laLeuAsnLysAspArgileuLysAlaValArgGluLysGlyValThrTyrLysG 288  
 QY 3078 GGAAGCCATCAGACTAACAGCGGATCCCGGAGAAACCTCAAGCCAGAGAGAGT 3137  
 Db 288 lyArgProIleArgileThrProAspPheSerProGluThrMetLysAlaArgAlat 308  
 QY 3138 GAGGCCAATATCCATCTTAAAGAAATATTTTCAACCCAGAAATTCATATCCAG 3197  
 Db 308 rpThrAspValIleGlnThrLeuArgGluHisLysCysGlnProArgLeuTyrProA 328  
 QY 3198 CCAACCAAGCTCTTAAGTGAAGGAGAGAAATCTCTACAGAGAAAGCAATGCTGA 3257  
 Db 328 laLysLeuSerIleThrIleAspGlyGluThrLysValPheHisAspLysThrLysPheT 348  
 QY 3258 CAGATTTTGTCCACACAGCCGCTCTTACAGAGCTCTCTGAGGAGAGCAACATGG 3317  
 Db 348 hrGln-TyrLeuSerThrAsnProAlaLeuGlnArgileThrGluLysLysGlnTyr 367  
 QY 3318 AAA 3320  
 Db 368 Lys 368

## RESULT 12

ALU6\_HUMAN STANDARD; PRT; 593 AA.

AC P39193; 1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SP sequence contamination warning entry.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M.; Makalowski W.;  
 RT "Alu alert."  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences."  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J.; Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 ..

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPEATIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U14572; ; NOT\_ANNOTATED\_CDS.  
 DR KW Hypothetical protein.  
 FT DOMAIN 1 97 FRAME-1.  
 FT DOMAIN 101 196 FRAME-2.  
 FT DOMAIN 200 295 FRAME-3.  
 FT DOMAIN 299 395 FRAME-4.  
 FT DOMAIN 399 494 FRAME-5.  
 FT DOMAIN 498 593 FRAME-6.  
 SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;  
 Alignment Scores:  
 Pred. No.: 1-39e-19 Length: 593  
 Score: 385.00 Matches: 80  
 Percent Similarity: 87.76% Conservative: 6  
 Best Local Similarity: 81.63% Mismatches: 9  
 Query Match: 0.73% Indels: 3  
 DB: 1 Gaps: 1  
 US-10-083-853B-2 (1-29921) x ALU6\_HUMAN (1-593)  
 QY 16853 GGCCGGGCGCAGTGGCTCAGCGCTGTAATCTAGCTTTGGAGGCCGAGTGGTGA 16794  
 Db 1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGlyGly 20  
 QY 16793 TCACCTGAGGTGAGGATTCAGAACACAGCGTGGCCACATGGCAAAACCCGTCCTACT 16734  
 Db 21 SerProGluValGlySerArgProAla\*\*\*ProThrTrpArgAsnProValSerThr 40  
 QY 16733 AAAATACAAAAATTAGCCAGGCGTGTATGGCAGGCACCTTAATCCAGCTACTCAGGAG 16674  
 Db 41 LysAsnThrLysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGlu 60  
 QY 16673 GCTGAGCGGGAGAGTACCTTGAACCCGGAGCGGAGGTGCGATGAGTCGATCGCG 16614  
 Db 61 AlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuArg\*\*\*AlaGluIleAla 80  
 QY 16613 CCATTGCACCTCCAGCTGGGTAAAA-----GAACTCCATCTCAAAAAAAA 16568  
 Db 81 ProLeuHisSerLeuGly-AsnLysSerGluThrProSerGlnLysLys 97

```
RESULT 13
ALU8_HUMAN
ID AC P39195; STANDARD; PRT; 591 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
FT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX', IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U14574; ; NOT_ANNOTATED_CDS.
DR Hypothetical protein.
KW DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:
Pred. No.: 2.49e-19 Length: 591
Score: 381.50 Matches: 79
Percent Similarity: 85.57% Conservative: 4
Best Local Similarity: 81.44% Mismatches: 10
Query Match: 0.73% Indels: 4
DB: Gaps: 1

US-10-083-853B-2 (1-29921) x ALU8_HUMAN (1-591)
QY 16954 AGCCGGGCGCAGTGGCTCAGCGCTGTAATCCAGCACTTTGGAGCGCGAGTGGGTGG 16795
D 4 ArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGly-Arg----- 20
QY 16794 ATCACTGAGTGTGAGGAGTTCACAGACAGCGTGGCCCAACATGCCAAACCCGCTCTCTAC 16735
D 21 -SerProGluValArgSerArgProAlaTrpProThrTrp***AsnProValSerTh 40
QY 16734 TAAAAATACAAAATTAAGCAGCGCTGATGGCAGGCACCTGTATATCCAGCTACTCAGGA 16675
D 40 rLysAsnThrLysIleSerArgAlaTrpArgAlaProValIleProAlaThrArgGl 60
QY 16674 GGCTGAGCGGGAGAAATCACTTGAACCCGGGAGCGAGGTTCAGTGCAGTGCAGTGCAGTGC 16615
D 60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgLeuGln***AlaGluIleAl 80
QY 16614 GCATTCCTCCAGCTGGTGGTAAAGAAACTCCATCTCAAAAAA 16566
D 80 aProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys 96

RESULT 14
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
FT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
```

RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U14571; -; NOT\_ANNOTATED\_CDS.  
 KW Hypothetical protein.  
 FT DOMAIN 1 95 FRAME-1.  
 FT DOMAIN 99 193 FRAME-2.  
 FT DOMAIN 197 291 FRAME-3.  
 FT DOMAIN 295 389 FRAME-4.  
 FT DOMAIN 393 487 FRAME-5.  
 FT DOMAIN 491 585 FRAME-6.  
 SQ SEQUENCE 585 AA; 63957 MW; 46BE8C4F493650A7 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.24e-17 Length: 585  
 Score: 358.00 Matches: 78  
 Percent Similarity: 75.65% Conservative: 9  
 Best Local Similarity: 67.83% Mismatches: 18  
 Query Match: 0.68% Indels: 10  
 DB: 1 Gaps: 2  
 US-10-083-853b-2 (1-29921) x ALU5\_HUMAN (1-585)  
 QY 16884 ATATAGACACCTTATTGACATAAAGAT-----AGGCGGGC 16846  
 Db ::::::::::: : : : :  
 179 LeuHisSerLeuAlaThrGluArgAspSerValSerLysLys\*\*\*\*\*ProGly 198  
 QY 16845 GCAGTGGCTCAGCCTGTAATCCTAGCAGCTTTGGAGCGCGAGGTGGGTGGATACCTGA 16786  
 Db ::::::::::: : : : :  
 199 AlavalAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIleThr--A 218

QY 16785 GGTCCAGGAGTTCACGAACACAGCGTGCCAAACATGGCAAAACCCCGTCTCTACTTAAATAATAC 16726  
 Db ::::::::::: : : : :  
 218 rgSerArgAspArgAspHisProGlyGlnHisGlyGluThrProSerLeuLeuLysIleG 238  
 QY 16725 AAAATTACCCAGCGGTGATGCGAGCGACCTGTAAATCCAGCTACTCAGGAGGCTCAGGC 16666  
 Db ::::::::::: : : : :  
 238 InLysLeuAlaGlyArgGlyGlyAlaArgLeu\*\*\*SerGlnLeuLeuGlyArgLeuArgG 258  
 QY 16665 GGAGAAATCACCTTGAACCCGGAGCGGAGGTTCAGCTGAGCTGAGATCGCGCATTCGA 16606  
 Db ::::::::::: : : : :  
 258 InGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysT 278  
 QY 16605 CTCACGCTGG---GTAAGAAAGAACTCCATCTCAAAAAAAA 16568  
 Db ::::::::::: : : : :  
 278 hrProAlaTrpArgGlnSerGluThrProSerGlnLysLys 291  
 RESULT 15  
 ALU5\_HUMAN  
 ID ALU5\_HUMAN STANDARD; PRT; 585 AA.  
 AC P39192;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SC sequence contamination warning entry.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=31384422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----

CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U14571; -; NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein.  
FT DOMAIN 1 95 FRAME-1.  
FT DOMAIN 99 193 FRAME-2.  
FT DOMAIN 197 291 FRAME-3.  
FT DOMAIN 235 389 FRAME-4.  
FT DOMAIN 393 487 FRAME-5.  
FT DOMAIN 491 585 FRAME-6.  
SQ SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:  
Pred. No.: 6e-17 Length: 585  
Score: 348.50 Matches: 75  
Percent Similarity: 78.35% Conservative: 1  
Best Local Similarity: 77.32% Mismatches: 18  
Query Match: 0.67% Indels: 3  
DB: .1 Gaps: 1  
US-10-083-853B-2 (1-29921) x ALU5\_HUMAN (1-585)  
QY 16569 TTTTGTGAGATGGAGTTCTTTT---ACCCAGGCTGGAGTCAATGGCGGATCTCAGC 16625  
Db 393 PhePhe\*\*\*AspGlyValSerLeuCysArgGlnAlaGlyValGlnTrpArgAspLeuGly 412  
QY 16626 TCACGTGCAACCTCCGCCCTCCCGGGTTCAGGTGATTCCTCCGCCCTCAGCTCTCTGAGTAGC 16685  
Db 413 SerLeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSer 432  
QY 16686 TGGGATTACAGTGCCTGCCATCAGCCTGCGCTAAATTTTGTATTTTAGTAGACACGGG 16745  
Db 433 TrpAspTyrArgArgAlaProArgProAlaAsnPheCysIlePheSerArgAspGly 452  
QY 16746 GTTTTGCATGTGGCCACGCTGTCTTGAACCTCCTGACCTCAGCTGATCCACCCACCTC 16805  
Db 453 ValSerProCysTrpProGlyTrpSerArgSerLeuAspLeu--ValIleArgProProA 472  
QY 16806 GGCCTCCCAAGTGTAGATTACAGGCTGAGCCACTGCGCCCGGC 16852  
Db 472 rgProProLysValLeuGlyLeuGlnAla\*\*\*AlaThrAlaProGly 487

Search completed: July 4, 2003, 20:08:09  
Job time : 462.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 14, 2003, 20:07:18 ; Search time 429 Seconds  
(without alignments)  
13409.968 Million cell updates/sec

Title: US-10-083-853B-2

Sequence: 1 grtatgtgaagaagctca.....caattctgtgaagaagtaa 29921

Scoring parameters: BLOSUM62  
Gapop 10.0, Xgapext 0.5  
Gapop 10.0, Ygapext 0.5  
Gapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10083853/runat\_03072003\_093612.8103/app\_query.fasta.1.30087  
-DB=PIR73 -QPMF=fastan -SUFFIX=pr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US10083853.ecgn.1.1.723.ernat\_03072003\_093612.8103 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: B34087  
1: pir2.\*  
2: pir3.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5336	10.2	1280	2	B34087
2	5195	9.9	1275	2	S65824
3	5169	9.9	1275	2	I38588
4	5139	9.8	1259	4	GNHUL1
5	5108	9.8	1275	2	B28096
6	3658.5	7.0	1260	4	GNLRL1
7	3552.5	6.8	1281	1	GNMSLL
8	2880.5	5.5	712	2	S23650
9	2257	4.3	1280	2	B34087
10	2229	4.3	1275	2	I38588
11	2215	4.2	1275	2	I38588
12	2196	4.2	1275	2	B28096
13	2183	4.2	712	2	S23650
14	2176.5	4.2	562	2	JU0033

C	15	2145	4.1	1259	4	GNHUL1	retrovirus-related
	16	1932	3.7	685	2	S16783	probable RNA-direc
	17	1632.5	3.1	556	2	S21347	hypothetical prote
C	18	1479.5	2.8	1260	4	GNLRL1	retrovirus-related
	19	1466.5	2.8	513	2	S21976	probable RNA-direc
C	20	1450.5	2.8	1281	1	GNMSLL	retrovirus-related
	21	1190.5	2.3	500	2	S16788	probable reverse t
	22	1119.5	2.1	1258	2	T14853	reverse transcript
C	23	1117	2.1	562	2	JU0033	hypothetical l1 pr
	24	1104.5	2.1	1258	2	T14855	reverse transcript
	25	1090.5	2.1	338	2	I38587	retrovirus-related
C	26	995.5	1.9	275	2	S21348	probable pol polyp
	27	972	1.9	335	4	A26718	retrovirus-related
	28	969.5	1.9	340	2	PH0217	reverse transcript
C	29	962.5	1.8	392	2	S57662	hypothetical prote
	30	850	1.6	275	2	S21348	probable pol polyp
	31	826.5	1.6	392	2	S57662	hypothetical prote
C	32	746	1.4	513	2	S21976	probable RNA-direc
	33	704	1.3	143	2	E41925	hypothetical prote
C	34	636	1.2	143	2	E41925	hypothetical prote
	35	485.5	0.9	138	2	I59767	lamin-like protein
	36	463.5	0.9	202	2	S21346	probable pol polyp
	37	443	0.8	1308	2	B32494	transposable eleme
	38	439.5	0.8	190	2	B25556	hypothetical prote
	39	422.5	0.8	1152	2	S20106	hypothetical prote
	40	416	0.8	351	2	S21345	retrovirus-related
	41	415.5	0.8	1231	2	C84716	hypothetical prote
C	42	409	0.8	370	2	JC7591	spinal cord-derive
	43	409	0.8	762	2	T47440	hypothetical prote
	44	399.5	0.8	872	2	H86435	protein F17F8.5 [1
	45	399	0.8	1344	2	H84557	hypothetical prote

ALIGNMENTS

RESULT 1

B34087  
hypothetical protein (L1H 3' region) - human  
C:Species: Homo sapiens (man)  
C>Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 30-Sep-1993  
C:Accession: B34087  
R:Scott, A.F.; Schmeckpeper, B.J.; Abdelrazik, M.; Conney, C.T.; O'Hara, B.; Rossiter, Genomics 1, 113-125, 1987  
A:Title: Origin of the human L1 elements: proposed progenitor genes deduced from a co A:Reference number: A34087; MUID:88085185; PMID:3692483  
A:Accession: B34087  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1280 <SCO>  
C:Superfamily: pol polyprotein

Alignment Scores:  
Pred. No.: 0  
Score: 5336.00  
Percent Similarity: 79.94%  
Best Local Similarity: 75.23%  
Query Match: 10.20%  
DB: 2  
Length: 1280  
Matches: 1084  
Conservative: 68  
Mismatch: 123  
Indels: 173  
Gaps: 7

US-10-083-853B-2 (1-29921) x B34087 (1-1280)

QY	3413	CCAGCTAGTGTCTAATGGCAGGATCAATATTCACACATATATATTAACCTTAATGTA	3472
DB	1	ProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnVal	20
QY	3473	ATFGGCTAAATTCCTCCCAATTAAGACACAGACTGGCAATTAAGAGTCAAGAC	3532
DB	21	AsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAsp	40
QY	3533	CCATCAGTGTGCTGTTATTCAGGAGGCCCATCTCACATGAACACACATAGGCTCAA	3592
DB	41	ProSerValCysCysIleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLys	60



```
QY 3593 ATAAAGGGATGGAGGAAGATTACCAAGTAATGGAAAAACAAAAAGCAGGGTT 3652
Db 61 IleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyVal 79
QY 3653 GCAATCCTAGTCTCTGATATAACAGACTTTAAACCAACAAGATCAAAAGACAAAGAA 3712
Db 80 AlaIleuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGlu 99
QY 3713 GGCCATTACATATGTAAGGCATCAATGGCAAGAGAGCTAACTATCTTAATATA 3772
Db 100 GlyHisTyrIleMetValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIle 119
QY 3773 CATGCACCAATACAGAGACCCAGATTCAATAAGCAAGTCTTTAGAGACCTACAAGA 3832
Db 120 TyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArg 139
QY 3833 GACTTTGACTCCACACAATACTAGTGGGAGTCTAAATAATAATAGACACTTTACACC 3892
Db 140 AspLeuAspSerHisThrIleIleMetGly-----AspPheAsnThr 153
QY 3893 CCACCTGCCAATATTAGCGAGATCAATGACAGAGAAAATTAACAAGGATATCCAGAGTTG 3952
Db 154 ProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAspIleGlnGluLeu 173
QY 3953 AACTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCAACCCCAATCA 4012
Db 174 AsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSer 193
QY 4013 ACAGATATACACTCTCTCAGCATCACATACACTATTTTAAATTCACCATTAATT 4072
Db 194 ThrGluTyrPhePheSerAlaProHisHisThrTyrSerLysIleAspHisIleLeu 213
QY 4073 TTAAGTAAACACTCTCTCAGCAATGCAAAAGAACAGAAATCTTAACAACAGTCTCTCA 4132
Db 214 GlySerLysAlaLeuLeuSerLysCysLysArgThrGluIleIleThrAsnCysLeuSer 233
QY 4133 GACTACAGTGCATCTATTAGACTCAGAAATTAGAAACTCAGCTCAAAATCACACAAT 4192
Db 234 AspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnHisSerThr 253
QY 4193 ACATGAAACTGAACAACCTGCTCTCGAATGACTACTGGTAAATAACAAAATGAAGCA 4252
Db 254 ThrTpyLysLeuAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAla 273
QY 4253 AAAATAAGATGTCTTTGAAACCAATGAGAACAAAGACACAATGTACCAGAAATCTCTGG 4312
Db 274 GluIleLysAsnPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrp 293
QY 4313 GGCATATTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCTACAGAGAAG 4372
Db 294 AspThrPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLys 313
QY 4373 CAGGAAATATCTAAATACACACTTAACATCACAAATTAAGAAGTACAGAGAAGAG 4432
Db 314 GlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGlu 333
QY 4433 CAAACAAATTCAAAAGCTAGCAGACAGAAAGAAATAACTAAGATCAGACGAACTGAAG 4492
Db 334 GlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaGluLeuLys 353
QY 4493 GAGATAGACACAAAAAGCCCTTCAAAATAATCAATGAATCCAGAGCTGGTTTTTGA 4552
Db 354 GluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGlu 373
QY 4553 AAAGATCAGCAAAAT---AGACCACCTAGACAGACTAATAAGAGAAAAGAGAGAANA 4608
Db 373 uLysIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysGlyGluLysAs 393
QY 4609 TCAAGAGATGCAATAAAAAATGATAAAGGGGATATCACCCGATCCACAGAAATACA 4668
Db 393 nGlnIleAspSerIleLysAsnAspLysGlyAspIleThrAspProThrGluIleGlu 413
QY 4669 AACTATTATCAGAGAAATATTATAAACACCTCTATGCAAAATAAAGTAGAAAATCTAGAAGA 4728
```

```
Db 413 nThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGlu 433
QY 4729 AATGGATAAATCTCTGGGACACATATGTAGCTGCTATGGACCTTGGGGGACACACAAG 4788
Db 433 uMetAspLysPheLeuAsp----- 439
QY 4789 GGGGTCAATGCAGAAATAAAAGACAAAGACAAAAGAGATGTGTTTGAAGTAGGGGTCAAG 4848
Db 439 ----- 439
QY 4849 GGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCACCCCTCTGTATTATTA 4908
Db 439 ----- 439
QY 4909 GCAAAAGAGATAGCGAGAGGGTGAGTTGGAAGAGAGGTGCTAGCTGTTAGGTCCAGAGTA 4968
Db 439 ----- 439
QY 4969 GGCCTGCAAGACTGCATTCTCAAAACAATAGGCTCTAGATGTGCCAGTAGATAACCTCAA 5028
Db 439 ----- 439
QY 5029 GGAGCCAGTGCAGGGAGTGATGGCCCTCAGCAAAACCTTCTTAGGCGAGCACAGAGTAA 5088
Db 439 ----- 439
QY 5089 GTTTCGCCACATTCGTATTTCAGATAAACAGTTTGTGCTTTGTGATCAAGTAGCCTCCAGT 5148
Db 439 ----- 439
QY 5149 GGAATGCTGAGTTGCTGATGATCCCTTTGGCCTTTTGGCTCCCAAAAACACATACACCT 5208
Db 440 -----ThrTyrThrLe 443
QY 5209 CTCACAGACTAAACAGGAGAGAGTCAAAATCCCTGAATATACCAAGTAAACAGTTCTAAAT 5268
Db 443 uProArgLeuAsnGlnGluGluValGluSerLeuAsnArgProIleThrGlySerGluIle 463
QY 5269 TGAACGAGTAAATTGATAGCTTACCAACCAAAAAAGTCCAGGACACAGCGGATTCCACAGC 5328
Db 463 eGluAlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrAla 483
QY 5329 CAATTTCTACAGAGGTACAAAGAGAAGCTGGTACTATTCTCTTGAACACTATTCCAAA 5388
Db 483 aGluPheTyrGlnArgTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSe 503
QY 5389 AATAGAA---AATGGGAATCCCTCACTCACTTTTACGAGCCAGCATCATCTGTATAC 5445
Db 503 rIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleP 523
QY 5446 CAAAACCTAGCAGTGACACACAACAAAAGAGGAAATTTTCAGGCCCATATCCCTGTATGAACA 5505
Db 523 rOlySProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnI 543
QY 5506 TTGATGTGAAAATCCCTCAATAAAATACCTGGCAAAACCAATCCAGCAGCACATCAAAAAGC 5565
Db 543 leAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGlnHisIleLysLysL 563
QY 5566 TTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATGCAAGGCTGGTTCAAAATATATCA 5625
Db 563 euIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgL 583
QY 5626 AATCAATAAATCTAGGCCCATCACATAAAACAGAACCAATGACAAAAACACATGATTATCT 5685
Db 583 ySerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleS 603
QY 5686 CAATAGATGCAGAAAGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAAATAATCTCA 5745
Db 603 erIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuA 623
QY 5746 GTAACACTAGGTATCGATGGAATGTATCTCAAAATAATAAGAGCTATTTATATAC-AAACCCA 5804
Db ----- 5804
```



Alignment Scores:

Pred. No.:	0	Length:	1275
Score:	5195.00	Matches:	1057
Percent Similarity:	78.97%	Conservative:	77
Best Local Similarity:	73.61%	Mismatches:	136
Query Match:	9.93%	Indels:	173
DB:	2	Gaps:	7

US-10-083-853B-2 (1-29921) x S65824 (1-1275)

3428	ATGCGAGGATCAAAATTCACACATATATATTAATCTTAATGTAAATGGCTAAATTC	3487
QY		
1	MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer	20
Db		
3488	CCAAATTAAGACACAGACTGGCAAAATTGGATAAAGAGTCAACAGCCCATCAGTGTGCTGT	3547
QY		
21	AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys	40
Db		
3548	ATTGAGGAGCCCATCTCACATGAAAGACACACATAGGCTCAAAATAAAGGATGGAGG	3607
QY		
41	IleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTpArg	60
Db		
3608	AAGATTACCAAGTAATGGAAAAACAAAAAACCAGGGGTGGCAATCCCTAGTCTCT	3667
QY		
61	LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer	79
Db		
3668	GATAAAACAGACTTTAAACCAACAAAGATCAAAAGAGACAAAGAGCCATTACATAATG	3727
QY		
80	AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet	99
Db		
3728	GTAAGGCATCAATGGAACAAGAGAGCTTAACCTATCTTAATATACATGACCCCAATACA	3787
QY		
100	ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr	119
Db		
3788	GGAGCACCAGATTCAATAAGCAAGTCTTTAGAGACCTACAAAGAGACTTTTGACTCCAC	3847
QY		
120	GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis	139
Db		
3848	ACAATAATAGTGGAGTCTAAATAATAAATAGACACTTTAACACCCCACTGCCCAATATTA	3907
QY		
140	ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu	153
Db		
3908	GGCAGATCAATGAGACAGACAAAAATTACAAGGATATCCAGGAGTTGAACCTGAGCTCGAC	3967
QY		
154	AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis	173
Db		
3968	CAAGCGGACTAATAGATATCTACAGAACTCCCAACCCCAAAATCAACAGAAATATACACT	4027
QY		
174	GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe	193
Db		
4028	TTCTCAGACTCACATTACACCTATTTTAAATTTGACCATGTAAATTTAAGTAAACACTC	4087
QY		
194	PheSerAlaProHisIsthrTyrSerLysIleAspHisIleValGlySerLysAlaLeu	213
Db		
4088	CTCAGCAATGCAAAAGACAGAAATCTTAACAAACAGTCTCTCAGACTACAGTGCATC	4147
QY		
214	LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle	233
Db		
4148	TATTTAGAAGTCTAGAATTAAAGAACTCACTCAAAATCACACAACTACATGGAACCTGAAC	4207
QY		
234	LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAsn	253
Db		
4208	AACCTGCTCTGAATGACTACTGGGTAAATAACAAATAAGAGCGCAAAATAAGATGTTTC	4267
QY		
254	AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe	273
Db		
4268	TTTGAACCAATCAGAACAAAGACACAATGTACACGAATCTCTGGGGCATATTTAAGCA	4327
QY		
274	PheGluThrAsnGlnAsnLysAspThrThrTyrGlnAsnLeuItrPaspAlaPheLysAla	293
Db		
4328	GTGTGTAGGGGAAATTTTATAGCACTAGATGCCCTACAGAGAAAGCAGGAAATATCTAAA	4387
QY		

Db	294	ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys	313
Qy	4388	ATAGACACCTTAACATCACATAATTAAGAAGAACTAGAGAAGAGAGCAAAATTCAAAA	4447
Db	314	IleAspThrLeuThrSerGlnLeuLysGluLeuGlnGlnThrHisSerLys	333
Qy	4448	GCTAGCAGAGACACAAGAAATAACTAAGATCAGAGCAGAACTGAAGGAGATAGAGACAAA	4507
Db	334	AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln	353
Qy	4508	AAAGCCCTTCAATAAATCAATCAATCCAGGAGCTGGTTTTCGAAAAGATCAGCAAAAT	4567
Db	354	LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGluArgIleAsnLysII	373
Qy	4568	----AGACCCTAGACAGACTAATAAAGAAGAAAAGAGAGACAAGTCAAGAGATGCAAT	4623
Db	373	eAspArgProLeuSerArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrII	393
Qy	4624	AAAAAATGATAAGGGGATATCACCCGATCCCCAGAAAATAACAACTATTATCAGAGA	4683
Db	393	eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrIleArgGI	413
Qy	4684	ATATTATAACACCTCTATCCAAATAAACTAGAAAATCTAGAGAAAATGGATAAAATTCCT	4743
Db	413	uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspThrPheLe	433
Qy	4744	GGACACATATGTAGCCTGTATGGACCTTGGGGCAGACAGACAAAAGGGGTGAATGCAGAA	4803
Db	433	uAsp-----	434
Qy	4804	ATAAAGACAAAAGACAAAGAGTAGTTTGGAAAGTAGGGGTACAGGGGCAACTTGCCTCT	4863
Db	434	-----	434
Qy	4864	AATGGACAGGGCCCTGAGCTTTACACCCACCTCTGTATTATTAGCAAAAGAGATAGC	4923
Db	434	-----	434
Qy	4924	GAGAGGGTGAGTTGGAAGAGAGGTGAGCTGTAGTTCACAGATAGGCGCTGCAAGACTGC	4983
Db	434	-----	434
Qy	4984	ATTCCTCAAAATAAGGCTCTAGATGTGCCAGTAGATAACCTCAAGGAGCGAGTGCCAGG	5043
Db	434	-----	434
Qy	5044	GAGTGATGGCCCTCAGCAAACTTCTAGGCGCAGGCACAGAGTAAGTTTGCCCACTTCT	5103
Db	434	-----	434
Qy	5104	GTATTACGATAAACAGTTTGCTGTTTGATCAAGTAGCCTCCAGTGGAAATGCTGATTGG	5163
Db	434	-----	434
Qy	5164	TCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACCA	5223
Db	435	-----ThrTyrThrLeuProArgLeuAsnGI	443
Qy	5224	GGAGAAGTCAATCCCTGTAATATACAGTAACAGTTCTAAATTTGAAGCAGTAATTGA	5283
Db	443	nGluGluValGluSerLeuAsnArgProIleThrGlySerGluIleValIleIleAs	463
Qy	5284	TAGCCTACCAACCAAAAAGTCCAGACAGAGGATTCACAGCCAAAATTCCTACCAGAG	5343
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnArg	483
Qy	5344	GTACAAAGAGAAGCTGGTACTATTCTCTCTGAAACTATTCCAAAAAATAGAA---AATGG	5400
Db	483	gTyrMetGluGluLeuValProPheLeuLysLeuPheGlnSerIleGluLysGlu-G	503
Qy	5401	GAATCCTCCCTAACTCATTTTACGAGCCAGCATCATCTGATACCAAAACCTTAGCAGTG	5460
Db	503	lvIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArgA	523

QY	5461	ACACACAAAAAGAGGAAATTTTCAGGCCATATCCCTGATGAACATTCATGTCACAAATCC	5520	QY	6533	GTAAACAAAAGCTGCTACTGGTACCAACAAAACAGATATATAGACCAATGGACAGAAC	6592
Db	523	spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL	543	Db	883	ValThrLysThrAlaTrpTrpTyrGlnAsnArgAspIleAspGlnTrpAsnArgThr	902
QY	5521	TCAATAAATACTGCGAACCAAAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580	QY	6593	GAGACCTCAGAATTTAC-ACTGCCATCTACATCCATCTCATCTTTTGACAAAACCTGCACAA	6651
Db	543	euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisAspG	563	Db	903	GluProSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys	922
QY	5581	AAGTTGGCGTCATCCCTGGGATGCAAGGCTGTCACAAATATGCAAAATCAATAATGTAG	5640	QY	6652	AACAAAGCAATGGAAAAAGGATTCCCTTATTATAATGTGTGGAAAAAATCGGCTAGCC	6711
Db	563	InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI	583	Db	923	AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla	942
QY	5641	GCCATCACATAAACAGAACCATGATGACAAAAACACATGATTTATCTCAATAGATGCAGAA	5700	QY	6712	ATATCGAGAAAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAAGTTAACTCAAGA	6771
Db	583	leGlnHisIleAsnArgAlaAsnAspLysAsnHisMetIleIleSerIleAspAlaGluL	603	Db	943	IleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsnSerArg	962
QY	5701	AGCCCTTGTCAAAATTCACAGCCCTTCATCTCTCAATAATTCCTAGTAACTAGGTATCG	5760	QY	6772	TCAATTAAGACTTAAATATAGACATATAAACCAATAAAACCCA-GAAGAAACCTTAGCC	6830
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623	Db	963	TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly	982
QY	5761	ATGGGAATGPATCTCAAAATAAAGAGCTATTATTATAC-AAACCCACAGCCAATATCATAC	5819	QY	6831	AATACCATTCAGATATGGCAAGACTTTCATGACTTAAACACCAACCAAAAGCAATG	6890
Db	623	spGlyThrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsnIleIleL	643	Db	983	IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet	1002
QY	5820	TGAATGGGCAAAACTGGAAGCATTCCTCTTGGAGAACTGGCACAGCAAGGATGCCCTC	5879	QY	6891	GCACAAAAGCCAAAATAGACAAGTGGGATCTGATTAAACTATAGAGCTTCGCACAGCA	6950
Db	643	euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663	Db	1003	AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
QY	5880	TCTCACCCTCTATTCAGATACTATTGGAGTTCTTGGCCAGGGCAATCAGGCATAGAC	5939	QY	6951	AAAAAAACTGTCATCAGAGTCAACAAAGCACTACAGAACTGGGAGAAATTTTGCAT	7010
Db	663	euserProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL	683	Db	1023	-LysGluThrThrIleArgValAsnArgGlnProThrTrpGluLysIlePheAlaTh	1042
QY	5940	AAGAAATAAGGCTATTCAAAATAGAAAGAGGAAGTATATTGTCCTGTTGGAGATG	5999	QY	7011	CTATCGATCTGACAAAGGCTAATATCCAGAGATCTACGAAGAACCTTAAACAAATTTACAA	7070
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703	Db	1042	rTyrSerSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGluIleTyrI	1062
QY	6000	ACATGTTTGTATATTTAGAAAACCCCATCGTCTCAGGCCAANAACCTCTTAAGCTGATA	6059	QY	7071	GAAAAA---AACAACCCCGTCAAAATATGCGCAAGGATATGAGCAGACACTTCTCAA	7126
Db	703	spMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleS	723	Db	1062	siLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerI	1082
QY	6060	GCAACTTCAGAACTCTCAGCACACAAATCAATGTCGAAAAATCACAAAGCATTTCTAT	6119	QY	7127	AGACACATTTATGAGCCCAACAAACATATGAAAAAACCTCATCATCATTCGTGTTAG	7186
Db	723	erAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPheLeu	743	Db	1082	sgLyspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr	1102
QY	6120	ACGCCAATAATAGACAAACAGAGCCCAATCATGATGAGTCAACTCTCAATTCACAATTGCTA	6179	QY	7187	AGAAATGCAAAACAAAACCCACAGTGCATACCATCTCATGCTAGTTAGTAGTGGTATCAC	7246
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheValIleAla	763	Db	1102	gGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI	1122
QY	6180	CAAAGAGATAAATACCTAGGAATACAACTTACAAGGGACACGCTAGGAACCTTCAAGG	6239	QY	7247	TAAAAGTCAGGAACACAAATGCTGGAGAGATGTGGAGAAATAGGAAACACTTTTCCA	7306
Db	763	erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783	Db	1122	eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyIleGlyThrLeuLeuH	1142
QY	6240	AGAACTACAAACCACTGATCAGGAATAAGAGAGGACACAAACAATGGAAAAACATTC	6299	QY	7307	CTGTTGGTGGGAATGTAATTTAGTCAACCATTTGGGAAGACAGTCTGGAGATTCTCTAA	7366
Db	783	luAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP	803	Db	1142	sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr	1162
QY	6300	CATGCTCAGATAGTAAGAATCAT-----GAAAAATGCCATCTAGCCCAAGTAAAT	6352	QY	7367	GGATCTAGAACCCAGAAATATCATTTGACCCCAATCCCATCTACTGAGTATATACCCAAA	7426
Db	803	roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT	823	Db	1162	gasPLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProAs	1182
QY	6353	ATAGATTACGTACCCCCCATCAAGCTACCATTTGACTTTCTTCACAGAAATTTGAAAAA	6412	QY	7427	GGAATATAAATCATTTATTAATAAGACACATGACACATATGTTTATTTCAGCAGCTGAT	7486
Db	823	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843	Db	1182	nGluTyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaAlaLeuPh	1202
QY	6413	CAACTTTAAATTCATATGGACCAAAAAAGAGCCACAGACCAAGCAACATCTTAAGC	6472	QY	7487	CACATAGCAAGACTTGGAAACCAACCCCAAAATGTCCATCAGTCATGATAGACTGATAAGAA	7546
Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862	Db	1202	ethrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL	1222
QY	6473	AAAAAGACAAAGCTGAGGTATCATGCTTACCTGACTTAAACTACTATATAGCTACA	6532	QY	7547	AACATGGCACATATACACCATCAAAATCTATTCAGCCATATAAAG-GATGAGTTCATGTC	7605
Db	863	GlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyrTrpLysAlaThr	882	Db	1222	smetTrpHisIleTyrThrMetGluTyrTrpAlaAlaLysAsnAspGluPheIleSe	1242
Db				QY	7606	CTTTCAGAGATATGGATGAAGCTGGAACCACTCATCTCTCAGCAAACTAACACAAAGA	7665

```

Db      1242 rPheValGlyThrTrpMetLysLeuGluThrIleLeuSerLysLeuSerGlnGluG1 1262
Qy      7666 GAAAAACCAACACACATGTTCTCCTACTGTTGAAGTGGGAGT 7705
Db      1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 3
138588
reverse transcriptase homolog - human retrotransposon L1
N:Alternate names: ORF2 protein
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 26-Aug-1999
C:Accession: I38588
R:Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.
Nature Genet. 7, 143-148, 1994
A:Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q
A:Reference number: I38587; MUID:95004577; PMID:7920631
A:Accession: I38588
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <RES>
A:Cross-references: EMBL:U09116; NID:q483914; PIDN:AAB60345.1; PID:q483916
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5169.00 Matches: 1054
Percent Similarity: 78.90% Conservative: 79
Best Local Similarity: 73.40% Mismatches: 137
Query Match: 9.88% Indels: 173
DB: 2 Gaps: 7

US-10-083-853B-2 (1-29921) x I38588 (1-1275)
Qy      3428 ATGGCAGATCAATTCACACATATATATTAACCTTAATGTAATGGCTAAATCC 3487
Db      1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20
Qy      3488 CCAATTAAGACACAGACTGGCAAAATTTGGATAAGAGTCAAGACCCATCATGCTGCTGT 3547
Db      21 AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
Qy      3548 ATTGAGGAGCCCTCTACATGAAAGACACACATAGCTAGCTCAAAATAAGGGATGGAGG 3607
Db      41 IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
Qy      3608 AAGATTACCAAGTAATGGAAACAAACAAAAAAGCAGGGGTGCAATCTAGTCTCT 3667
Db      61 LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyAlaIleLeuValSer 79
Qy      3668 CATAAACAGACTTTAAACCAACAAAGATCAAAAGAGACAAAGAGCCATTACATAATG 3727
Db      80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet 99
Qy      3728 GTAAAGGATCAATGGAAACAGAGAGCTACTACTCTAAATATACATGACCCCAATACA 3787
Db      100 ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr 119
Qy      3788 GGACACCCAGATTCAATAGCAAGTCTTTAGACACCTACAAAGAGACTTTGACTCCAC 3847
Db      120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139
Qy      3848 ACAATAATAGTGGGAGTCTAAATAATAATAGACACTTTACACCCCTCCACTCCCAATTA 3907
Db      140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerIleLeu 153
Qy      3908 GGCAGATCAATGACAGACAAATTAACAGGATATCCAGGAGTTGAAGTCTGGAC 3967
Db      154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHis 173
Qy      3968 CAACGGACCTTAATAGATATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTC 4027

```

```

Db      174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193
Qy      4028 TTCTCAGCATCACATTACACCTATTTAAATATGACCATGTAATTTAAAGTAAACACATC 4087
Db      194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213
Qy      4088 CTCAGCAATGCAAAAGAACAGAAATCTCTAACAAACAGCTCTCTCAGACTACAGTGAATC 4147
Db      214 LeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaIle 233
Qy      4148 TATTAGACTCACAATTAAGAACTCCTCACTCAAAATCACAACTACATGAAATGAACTGAC 4207
Db      234 LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn 253
Qy      4208 AACCTGCTCTGAATGACTACTGCTGTAATAACAAATGAAGCAAAATAAAGATGTC 4267
Db      254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe 273
Qy      4368 TTTGAAACCAATGAGAACAAAGACACAAATGTACAGAAATCTCTGGGCATATTTAAAGCA 4327
Db      274 PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293
Qy      4328 GTGTGTAGAGGAAATTTATAGCCTACATGCTCTACAGAGCAAGAGCAAGAAATATCTAAA 4387
Db      294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313
Qy      4388 ATAGACACCTTAACATCACATTAATAAGAACTAGAGAAAGAGAGCAAAACAAATTCAAA 4447
Db      314 IleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGlnThrHisSerLys 333
Qy      4448 GCTAGCAGAGACAAAGAAATAACTAAGATCAGACAGCAACTGGAAGGAGATAGAGACAA 4507
Db      334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353
Qy      4508 AAAGCCCTTCAATTAATCAATGAATCCAGGAGCTGCTTTTTCACAAAGATCACCAAAAT 4567
Db      354 LysThrLeuLys - IleAsnGluSerArgSerTrpPhePheGluArgIleAsnLysIle 373
Qy      4568 ----AGACCCTAGACAGACTAATAAGAAAGAGAGAGAGAGAAATCAAGAGAGATGCAAT 4623
Db      373 eAspArgProLeuAlaArgLeuIleLysLysArgGluLysAsnGlnIleAspThrIle 393
Qy      4624 AAAAAATGATAAGGGATATCACCCCGATCCACAGAAATACAAACTATTATTCAGAGA 4683
Db      393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgG1 413
Qy      4684 ATATTATAACACCTCTATGCAATAAATAACTAGAAAAATCTAGAAAAATGGATAAATTCCT 4743
Db      413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspThrPheLe 433
Qy      4744 GGACACATATGTAGCCTGTATGGACCTTGGGGGACAGAAACAAAGGGGTGAATGCAGAA 4803
Db      433 uAsp----- 434
Qy      4804 ATAAAGACAAAGACAAAGAGTATGTTGGAGTAGGGTTCAGGGGGCAACTTCCTCT 4863
Db      434 ----- 434
Qy      4864 AATGGCAAGGGCCCTGAGCTTTACACCACCTCTCTGTTATTATTATTAGCAAAAGAGATAGC 4923
Db      434 ----- 434
Qy      4924 GAGAGGTGAGTTGGAAGAGAGGTCACTGTTAGGTCCAGAGTAGCCCTGCAAGACTGC 4983
Db      434 ----- 434
Qy      4984 ATTCCTCAACAATAGGCTCTAGATGTCCAGTAGATAACCTCAAGGAGCCAGTGCAGG 5043
Db      434 ----- 434
Qy      5044 GAGTGTATGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAAAGTAAGTTTGCCACATCT 5103
Db      434 ----- 434

```

QY	5104	GTATTACGATAAACAGTTGCTGCTTTGTGATCAAGTAGCTCCAGTGGAGTGG	5163	QY	6180	CAAGAGAAATAAATACCTAGGAATACAACTTACAGGAGACAGCTAGGAACCTCTCTCAAG	6239
Db	434	-----	434	Db	763	erLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG	783
QY	5164	TCATGATCCCTTTGGCTTTTGGCTGCCAAAAACACATACACCTCTCAAGACTAACCA	5223	QY	6240	AGAACTACAAACCACTGATCAAGAAATAAGAGAGAGACAAACAAATGGAACAAATTC	6299
Db	435	-----	435	Db	783	luasnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP	803
QY	5224	GAAGAGTCAAACTCCCTGAATATACCAAGTAAAGTTCCTAAATGAACCAATGTA	5283	QY	6300	CATGCTCACAGATAGTAAGATCAT-----GAAATATGCATCTCCCAAGTAATAT	6352
Db	443	ngluGluValGluSerLeuasnGlyProIleThrGlySerGluIleValAlaIleLeas	463	Db	803	rocysSerTrpValGlyArgIleasnIleValLysMetAlaIleLeuProLysValIle	823
QY	5284	TAGCTACCAACCAAAAAAGTCCAGGACAGACGAGATTACAGCCAAATTTACACAGAG	5343	QY	6353	ATAGATTCACTGTCTACCCCATCAAGTACCATTTCTTCTCACAGAATTGGAACAAA	6412
Db	463	nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnAr	483	Db	823	yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT	843
QY	5344	GTCAAGAGAGAGCTGTACTATTCTCTCTGTAACATATTCACAAAAATAGAA---AATGG	5400	QY	6413	CAACTTTAAATTCATATGAAACCAAAAAAGAGCCACAGAGCCAGACAACTTTAAGC	6472
Db	483	gTyrLysGluGluLeuValProPheLeuLysLeuPheGlnSerIleGluLysGlu-G	503	Db	843	hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer	862
QY	5401	GAATCTCCCTAACTCATTTTACGAGCCAGCATCTCTGATACCAAAACCTAGCAGTG	5460	QY	6473	AAAAAGAACAAAGCTGGAGGTATCATGCTACCTGACTTAAACTATATAAGGCTACA	6532
Db	503	lylleuProasnSerPheTyrGluAlaSerIlelleuIleProLysProGlyArgA	523	Db	863	GlnLysasnLysAlaGlyIleThrLeuProTyrPheLysLeuTyrLysAlaThr	882
QY	5461	ACACAACAAAGAGAAATTTCCAGGCCCATATCCCTGTGATGAAACATTTGTAATCC	5520	QY	6533	GTAAACAAAACCTGCTACTGTGTACCAAAACAGATATATAGACCAATGGAACAGACA	6592
Db	523	spThrThrLysLysGluasnPheArgProIleSerLeuMetasnIleAspAlaLysIle	543	Db	883	ValThrLysThrAlaTrpTyrTrpTyrGlnasnArgAspIleaspGlnTrpAsnArgThr	902
QY	5521	TCAATAAATCTGGCAACCAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATC	5580	QY	6593	GAGACCTCAGAAATATAC-ACTGCATCTCATCATCTGATCTTTTGACAAACCTGACAAA	6651
Db	543	euasnLysIleLeuAlaasnArgIleGlnGlnHisIleLysLysLeuIleHisAspG	563	Db	903	GluproSerGluIleMetProHisIleTyrAsnTyrLeuIlePheAspLysProGluLys	922
QY	5581	AAGTGGCGTCATCCCTGGATCGAGGCTGGTTCAAAATATGCAATCAATTAATCTAG	5640	QY	6652	AACAAGCAATGGAAAAAGGATTCCTATTATTAATAATGGTGTGGAAAAACTGGCTAGCC	6711
Db	563	lnValGlyPheIleProGlyMetGlnGlyTrpPheasnIleArgLysSerIleasnValI	583	Db	923	asnLysGlnTrpGlyLysAspSerLeuPheasnLysTrpCysTrpGluasnTrpLeuAla	942
QY	5641	GCCATCATATAAACAGAACCAATACAAAAACACATGATATCTCAATAGATGCAGAAA	5700	QY	6712	ATATGCAAGAAAGCTGAAACCTGGATCCCTCTTACACCTTATACAAAAGTTAACTCAAGA	6771
Db	583	leGlnHisIleasnArgAlaLysaspLysasnHisMetIleIleSerIleaspAlaGlu	603	Db	943	IleCysArgLysLeuLysLeuaspLeuPheLeuThrProTyrThrLysIleasnSerArg	962
QY	5701	AGGCTTTGTCAAAATTCACAGCCCTTCATGTCTAAATAATCTCAGTAAACTAGGTATCG	5760	QY	6772	TGAATTAAGACTTAATATATAAGACATAAAACCAATAAAACCCCA-GAAGAAAACCTAGGC	6830
Db	603	ysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA	623	Db	963	TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluasnLeuGly	982
QY	5761	ATGGAATGTATCTCAAAATATAGAGCTATTTATAC-AAACCCACAGCCCAATATCATAC	5819	QY	6831	AATACCATTCAGGATGGACATGGCGCAAGACTTCATGACTATAAACACCAACAAAGCAATG	6890
Db	623	spGlyThrTyrPheLysIleAlaIleTyrAspLysProThrAlaasnIleIle	643	Db	983	IleThrIleGlnaspIleGlyValGlyLysaspPheMetSerLysThrProLysAlaMet	1002
QY	5820	TGAATGGCAAAACCTGGAAGCATTCCTTTTGAACTGGCAACAGAGGATGCCCTC	5879	QY	6891	GCACAAAACCCCAAAATAGACAAAGTGGGATCTGATTAACCTATAGAGCTTCTGCACAGCA	6950
Db	643	euasnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL	663	Db	1003	AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla	1022
QY	5880	TCACACACTCTTATCAAGATACTATTGGAAGTCTCGCCAGGCAATCAGGCAATAGA	5939	QY	6951	AAAAAAACCTCATCAGAGTGAAACAGCAACCTACAGATGGGAGAAAATTTTGGCAAT	7010
Db	663	euSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGlu	683	Db	1023	-LysGluThrThrIleArgValasnArgGlnProThrThrTrpGluLysIlePheAlaTh	1042
QY	5940	AAGAAATAAGGCTATTCAATAGAAAGAGAGAGGAGTCATATTGCTCTGTTTGCAGATG	5999	QY	7011	CTATCGATCTGACAAAGGCTTAATATCCAGAGACTTACGAGAACTTAAACAAATTTACAA	7070
Db	683	ysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA	703	Db	1042	rTySerSerAspLysGlyLeuIleSerArgIleTyrAsnGlnLeuLysGlnIleTyrLy	1062
QY	6000	ACATGTTGTATATTTAGAAAAACCCCTCTCTCAGGCCCAAAACCTCTTAAGCTGATAA	6059	QY	7071	GAAGAAA-----AACAAACCCCTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAA	7126
Db	703	spMetIleValTyrLeuGluasnProIleValSerAlaGlnasnLeuLysLeuLys	723	Db	1062	sLysLysThrAsnAsnProIleLysLysTrpAlaLysaspMetasnArgHisPheSerLy	1082
QY	6060	GCAACTTCAGAAAGTCTCAGGACACAAAAATCAATGTGCAAAAATCAACAGCATCTTAT	6119	QY	7127	AGAAGACATTTATGCAGCCCAACAAACATATGAAAAAACCTCATCATCTGGTGGTAG	7186
Db	723	erasnPheSerLysValSerGlyTyrLysIleasnValGlnLysSerGlnAlaPheLeuT	743	Db	1082	sGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSerSerLeuAlaIleAr	1102
QY	6120	ACCCCAATATACACAAACAGAGCCCAATATCATGAGTGAACCTCTCATTTCAATGCTA	6179	QY	7187	AGAAATGCAAAACCAACACAGTACATACCATCTCTCATCTAGTAGTGGTGGTACAC	7246
Db	743	yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlas	763	Db	1102	gGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleI	1122
				QY	7247	TAAAAAGTCAGGAACCAACAAATGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTCCA	7306



```
Db      1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHi 1142
QY      7307 CTGTTGGTGGGAATGTAATAGTTCACCACTGTGGAACAGAGTGTGAGATTCTTAA 7366
Db      1142 scYstrTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY      7367 GGATCTAGAACACAGAAATATCATTTTGACCCAGCAATCCCATTAAGTCTAGTATATACCAAA 7426
Db      1162 gAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTrpProG 1182
QY      7427 GGAATATAATCATTTATATAAGACACATGACACATATATGTTTATTCAGCAGCTGAT 7486
Db      1182 uAspTrpLysSerCysCysTrpLysAspThrCysThrArgMetPheIleAlaAlaLeuPh 1202
QY      7487 CACAATAGCAAGACTTGAACCAACCCAAATGTCATCAGTCATGATAGATGGAATAAGAA 7546
Db      1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL 1222
QY      7547 AACATGGCACATATACACCATGAATATATGACGCATATAAAG-GATGAGTTTCATGTC 7605
Db      1222 sMetTrpHisIleTrpThrMetGluTrpTyrAlaAlaIleLysAsnAspGluPheIleSe 1242
QY      7606 CTTTGCAGAGATATGATGAAGCTGGAACCATCATCTCAGCAAACTAACACAGAAACA 7665
Db      1242 rPheValGlyThrTrpMetLysLeuGluThrIleLeuSerLysLeuSerGlnGluGl 1262
QY      7666 GAAACCAACCAACCATCTCTCACTGCTGAAGTGGGAGT 7705
Db      1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 4
GNHUL1
retrovirus-related reverse transcriptase pseudogene - human
C:Species: Homo sapiens (nan)
C:Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 14-May-1999
C:Accession: A25313
R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.
Nature 321, 625-628, 1986
A:Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence
A:Reference number: A93381; MUID:86230917; PMID:2423883
A:Accession: A25313
A>Status: conceptual translation of pseudogene
A:Molecule type: DNA
A:Residues: 1-1259 <HAT>
A>Note: this sequence was constructed from an alignment of published and unpublished seq
C:Keywords: reverse transcriptase; pseudogene

Alignment Scores:
Pred. No.: 0 Length: 1259
Score: 5139.00 Matches: 1050
Percent Similarity: 79.00% Conservative: 71
Best Local Similarity: 74.00% Mismatches: 132
Query Match: 9.82% Indels: 173
DB: 4 Gaps: 7

US-10-083-853B-2 (1-29921) x GNHUL1 (1-1259)
QY      3434 GGATCAAAATTCACACATATATATTAACCTTAAATGGGCTAAATCCCAAT 3493
Db      2 GlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnAlaProIle 21
QY      3494 AAAAGACACAGACTGGCAATTTGGATAAGAGCTCAAGACCCCATCAGTGTGCTGATTTCAG 3553
Db      22 LysArgHisArgLeuAlaAsnTrpIleLysSerGlnAspProSerValCysCysIleGln 41
QY      3554 GAGGCCATCTCCATGAAAGACACATAGCTCAAAATAAAGGATGGAGGAAGATT 3613
Db      42 GluThrHisLeuThrCysArgAspThrHisArgLeuLysIleLysGlyTrpArgAsnIle 61
QY      3614 TACCAAGTAAATGGAACCAACCAAAAGAGGAGGTTGCAATCTAGTCTCTGATAAA 3673
Db      62 TyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSerAspLys 80
```

```
QY      3674 ACAGACTTTAAACCAACAAAGATCAAAAGAGACAAAGAGCCATTACATAATGTTAAAG 3733
Db      81 ThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTrpIleMetValLys 100
QY      3734 GCATCAATGGACACAGAAGAGCTAATCTCTTAATATATACATGACCCCAATACAGGACA 3793
Db      101 GlySerIleGlnGlnGluLeuThrIleLeuAsnIleTrpAlaProAsnThrGlyAla 120
QY      3794 CCCAGATTCATAAACAAGTCTCTTAGAGACCTTACAAAGAGACTTTTACCTCCACACAATA 3853
Db      121 ProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHisThrIle 140
QY      3854 ATAGTGGGAGTCTAAATAATAATAGACACTTTAACACCCCACTTACCAATATATAGGCAGA 3913
Db      141 IleMetGly-----AspPheAsnThrProLeuSerThrLeuAspArg 154
QY      3914 TCAATGAGACAGAAATTAACAAGGATATCCAGGAGTTGAAGTCTGAGCAACGCG 3973
Db      155 SerThrArgGlnLysIleAsnLysAspIleGlnGluLeuAsnSerAlaLeuHisGlnAla 174
QY      3974 GACCTAATAGATATCTACAGAACTCCCAACCCCAATCAACAGAAATATACACTCTCTCTCA 4033
Db      175 AspLeuIleAspIleTrpArgThrLeuHisProLysSerThrGluTrpThrPhePheSer 194
QY      4034 GCATCACATTACACCTATTTTAAATTTGACCATGTAAATTTTAAAGTAAAAACACTCTCTCAGC 4093
Db      195 AlaProHisHisThrTrpSerLysThrAspHisIleLeuGlySerLysThrLeuLeuSer 214
QY      4094 AAATGCAAAAGAACAGAAATCTCAACAACTCTCTCAGACTCAGTCACTGCAATCTATTATTA 4153
Db      215 LysCysLysArgThrGluIleIleThrAsnCysLeuSerAspHisSerAlaIleLysLeu 234
QY      4154 GAACCTCAGAAATTAAGAACTCACTCAAAATCACAACTACATGGAACCTGAACACCTG 4213
Db      235 GluLeuArgGlyLysLeuThrGlnAsnHisSerThrTrpLysLeuAsnAsnLeu 254
QY      4214 CTCCTGAATGACTACTGGGTAATAAACAAATGAAGCCAAATAAAGATGTTCTTTGAA 4273
Db      255 LeuLeuAsnAspTrpValHisAsnGluMetLysAlaGluIleLysLysPheGlu 274
QY      4274 ACCAATGAGACAAAGACACAAATGTACCAGATCTCTGGGCGATATTTAAACAGCTGT 4333
Db      275 ThrAsnGluAsnLysAspThrTrpGlnAsnLeuTrpAspThrAlaLysAlaValCys 294
QY      4334 ACAGGAAATTTATAGCTAGCTAGCTACAAGAGAAAGCAGCAATATCTATAATAGAC 4393
Db      295 ArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIleAsp 314
QY      4394 ACCTTAACATCAATTAAGAACTAGAGAGAAGAAAGACAAACAAATTCAAAAGCTAGC 4453
Db      315 ThrLeuIleSerGlnLeuLysGluLeuGluLysGlnGluThrAsnSerLysAlaSer 334
QY      4454 AGAAGACAAAGAAATAACTAAGATCAGACAGCAACTGAGGAGATAGACACAAAGAAC 4513
Db      335 ArgArgGlnGluIleIleLysIleArgAlaGluLeuLysGluIleGluThrGlnLysThr 354
QY      4514 CTTCAATAATAATCAATGAATCCAGGAGCTGGTGTGTTTGAAGAAAGATCAGCAAAAT 4569
Db      355 LeuGlnLys-IleAsnGluSerArgSerTrpPheGluLysIleAsnLysIleAspArg 374
QY      4570 ACCACTAGACAGCTAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4629
Db      374 gProLeuAlaArgLeuIleLysLysArgGluLysAsnGlnIleAspThrIleLysAs 394
QY      4630 TGATAAAGGGGATATCACACCGGATCCACAGAAATACAACTATATCAGAGAGATATTA 4689
Db      394 nAspArgGlyAspIleThrThrAspProThrGluIleGlnThrIleArgGluTrp 414
QY      4690 TAAACACCTCTATGCAAAATAACTAGAAATCTAGAGAAATGGAATAATCTCTCGACAC 4749
Db      414 rLysHisLeuTrpAlaAsnLysLeuGluAsnLeuGluMetAspLysPheLeuAsp--- 433
```





```
Db 1004 LysValLysIleAspArgTrpAspLeuIleLysLeuLysSerPheCysThrAla-LysGI 1023
QY AACTGTCATCAGAGTGAACAAGCAACCTACAGAAATGGGAGAAATTTTGGCAATCATCG 7016
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 uThrThrIleArgValAsnArgGlnProThrLysTrpGluLysIlePheAlaIleIyrse 1043
QY 7017 ATCTGACAAGGCTAATATCCAGAGATCTACGAGAACTTAAACAATTTACAGAAAAA 7076
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 rSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLysGlnIleTyrLysLysly 1063
QY 7077 ----RACACCCCGTCAAAATATGGCAAGGATATACGACACACTTCTCAAAAGAAGA 7132
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 sThrAsnAnProIleLysTrpAlaLysAspMetAsnArgHisProSerLysGluAs 1083
QY 7133 CATTATGACGCCCAACAACATATGAAAAAACCCTCATCATCTTGGTGGTAGAGAAAT 7192
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1083 pIleTyrAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleIleArgGluMe 1103
QY 7193 GCAAAACAACACAGATGACATACCATCTCATGCTAGTAGAATGGTGGATCACTAAAAA 7252
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1103 tGlnIleLysThrThrMetArgTyrHisLeuThrProValArgMetAlaIleIleLysLy 1123
QY 7253 GTCAGAGAAACAACAAATGCTGGAGAGATCTGGAGAAATAGGAACACTTTCCACHTTG 7312
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 sSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTr 1143
QY 7313 GTGGGAATGTAATATTAGTTCAACCATTTGGAGACAGCTGTGGAGATTCCTTAAGGATCT 7372
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 pTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLe 1163
QY 7373 AGAACCAAAATATCATTTGACCCCAATCCCATTTACTGAGTATATATACCCAAAGGAATA 7432
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 uGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProLysAspTy 1183
QY 7433 TAAATCATCTATTATAAGACACATGCACACATATGTTTATGTCAGCACTGATCAAT 7492
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1183 rLysSerCysCysTyrLysAspThrCysThrArgMetPheIleAlaAlaLeuPheThrIl 1203
QY 7493 AGCAAGACTTGGAAACCAACCCAAATGTCATCAGTATAGATGATGATAAGAAACATG 7552
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1203 eAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLysMetTr 1223
QY 7553 GCACATATACACCATGAATATATCCAGCCATATAAAG-GATGAGTTTCATGTCCTTGC 7611
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 pHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAspGluProMetSerProVa 1243
QY 7612 AGAGATATGGTGAAGCTGGAACCATCATCTCAGCAAACTAACACAA 7660
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1243 lGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGln 1259

RESULT 5
B28096 line-1 protein ORF2 - human
C:Species: Homo sapiens (man)
C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 30-Sep-1993
C:Accession: B28096
R:Skowronski, J.; Fanning, T.G.; Singer, M.F.
Mol. Cell. Biol. 8, 1385-1397, 1988
A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.
A:Reference number: A28096; MUID:88246405; PMID:2454389
A:Accession: B28096
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
C:Residues: 1-1275 <SKO>
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 0 Length: 1275
Score: 5108.00 Matches: 1045
Percent Similarity: 78.64% Conservative: 85
Best Local Similarity: 72.72% Mismatches: 140
Query Match: 9.76% Indels: 175
DB: 2 Gaps: 7
```

```
US-10-083-853B-2 (1-29921) x B28096 (1-1275)
QY 3428 ATGGCAGGATCAAAATTCACACATAATAATTAACCTTAATGCTTAATGGCTAAATTC 3487
Db 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnValAsnGlyLeuAsnSer 20
QY 3488 CCAATTTAAAGACACAGACTGGCAATTTGGATTAAGAGTCAAGACCCATCAGTGTCTGT 3547
Db 21 ProIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40
QY 3548 ATTGAGGAGGCCATCTCATGAAAGACACACATAGGCTCAAAATAAAGGATGAGG 3607
Db 41 IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60
QY 3608 AGATTTCACCAAGTAATGGAACAAAAAAGAGGGGTTTCCAATCTCAGTCTCT 3667
Db 61 LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79
QY 3668 GATAAACAGACTTTAAACCAACAAGATCAAAAGAGACAAAGAGCCCATACATAATG 3727
Db 80 AspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrMetMet 99
QY 3728 GTAAGGCAATCAATGGAACAAGAGCTAATCTTAAATATATACATGCACCCATACA 3787
Db 100 ValLysGlySerIleGlnGluLeuThrIleLeuAsnMetTyrAlaProAsnThr 119
QY 3788 GGAGCACCAGATTCTAAAGCAAGTTCTTAGAGACTTACAAAGAGACTTTGACTCCCAC 3847
Db 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuArgAspLeuAspSerHis 139
QY 3848 ACAATAATAGTGGGAGTCTTAAATAATAATAGACACTTTTAACACCCCACTGCC-AAATAT 3906
Db 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerThrLeu 153
QY 3907 AGCAGATCAATGAGACAGAAAAATTAACAAGATATCCAGAGTTGAACTGAGCTCTGGA 3966
Db 154 ArgGlnIleAsnGluThrGluSerGlnGlnGlyTyr-GlnGluLeuAsnSerAlaLeuH1 173
QY 3967 CCAAGGGACCTTAATAGATATCTACAGAATCCCCACCCCAATCAACAGATATACACT 4026
Db 173 sGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPh 193
QY 4027 CTCTCAGCATCACATTACACCTATTTTAAATTTGACCATGTAATTTTAAAGTAAACACT 4086
Db 193 ePheSerAlaProHisHisSerTyrSerLysIleAspHisIleLeuGlySerGluAlaLe 213
QY 4087 CCTCAGCAATGCAAAAGAACAGAAATCCTACAAACACTCTCTCAGACTCAGTCAAT 4146
Db 213 uLeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHisSerAlaI1 233
QY 4147 CTATTTAGAACTCAGAAATTAAGAACTCACTCAAAATATCAACAACATACATGGAACCTGAA 4206
Db 233 eLysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrThrTrpLysLeuAs 253
QY 4207 CAACCTGCTCTCCTCAATGACTACTGGTGAATAATAACAAAATGAAGGCAAAAATAAAGATGT 4266
Db 253 nAsnLeuLeuLeuAsnAspTy-TripValHisAsnGluMetLysAlaGluIleLysMetPh 273
QY 4267 CTTTGAACCAATGAGAACAAAGACACAATGTACCAAGATCTCTGGGGCATATTTAAAGC 4326
Db 273 ePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAl 293
QY 4327 AGTGTGTAGAGGAAATTTATAGCACTAGATGCCTACAGAGAAAGACAGAAATATCTAA 4386
Db 293 aValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLy 313
QY 4387 AATAGACACTTTAATCATCAAAATTAAGAAGCTAGAGAAAGAGCAACAAATTCAAA 4446
Db 313 sIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluThrHisSerLy 333
QY 4447 ACCTAGCAGAACAGAAATAACTAAGATCAGACGAGAACTGAAGGATAGACACACA 4506
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



Qy	6651	ANACAACGCAATGGAAAAAGGATTCCTTATTTAATAAATGGTGTGGAAAAAAGCTGGCTAGC	6711
Db	922	sasNlYsGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAl	942
Qy	6711	CATATGCAGAAAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAGCTTAACCTCAAG	6770
Db	942	alleCyseargLysLeuLysValaspProPheLeuThrProTyThrLysIleasnSerAr	962
Qy	6771	ATGAATTAAAGACTTAAATATAGACATATAAAACCATAAAAACCCA - GAAGAAAACCTTAGG	6829
Db	962	gTpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGl	982
Qy	6830	CAATPACCATTCAGGATATGACATGGCGCAAGAGCTTCATGACTATAAACACCAAAAGCAAT	6889
Db	982	ylleThrIleGlnaspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMe	1002
Qy	6890	GGCAACAAAAGCCAAAATAGACAAGTGGGATCTGATTAAACTATAGAGCTTCGCACAGC	6949
Db	1002	talaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAl	1022
Qy	6950	AAAAAAAACCTGCATCAGAGTGACCAAGCAACCTACAGAATGGGAGAAAATTTTGCAC	7009
Db	1022	a - LysGluThrThrIleargValasnArgGlnProThrThrTrpGluThrIlePheThrT	1042
Qy	7010	TCATTCGATCTGCACAAAGGCTAATATCCAGAGATCTACGAAGAACTTAAACAAATTTACA	7069
Db	1042	hrTySerSerAspLysGlyLeuIleSerArgIleTyAsnGluLeuLysGlnIleTyL	1062
Qy	7070	AGAAAA - ---AACACCCCGTCNAATATGGCAAGGATATGACGACAGACTTCTCAA	7125
Db	1062	ysLysLysThrAsnAspProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL	1082
Qy	7126	AAGNAGACATTTATGCAGCCCAACACATATGAAAAAACCTCATCATTCATTTGGTCTTA	7185
Db	1082	ysGluaspIleTyAlaAlaLysHisMetLysLysCysSerSerSerLeuAlaIleA	1102
Qy	7186	GAGAAATGCAAAACCAACACAGTCACATACCATCTCATGCTAGTTAGAAATGGTGATCA	7245
Db	1102	rgLluMetGlnIleLysThrThrMetArgTyHisLeuThrProValArgMetAlaIleI	1122
Qy	7246	CTAAAAAGTCAGAAACAAACAAATGCTGGAGAGAGTGGAGAAATAGGAACACTTTTCC	7305
Db	1122	leLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuH	1142
Qy	7306	ACTGTTGGTGGGAATGTAATTAGTTCAACCATTTGGAAACAGAGTGGAGAGTTCCTTA	7365
Db	1142	iscysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuA	1162
Qy	7366	AGGATPAGAACCCAGAAATATCATTTGACCCGCAATCCCATTTACTGAGTATATACCCAA	7425
Db	1162	rgAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProA	1182
Qy	7426	AGGAATATAATCTATTTATAAGACACATGCACACATATGTTTATTTGACGACTGA	7485
Db	1182	snaspTyLysSerCysCysTyLysAspThrCysThrArgMetPheIleAlaLeuP	1202
Qy	7486	TCACAATAGCAAAAGACTTGAACCAACCCAAATGTCATCAGTGATAGACTGGATAAAGA	7545
Db	1202	hethrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysL	1222
Qy	7546	AAACATGGCACATATACACCATGAATATCATTCATCGAGCCATAAAAAG - GATGAGTTCTAGT	7604
Db	1222	ysMetTrpHisIleTyThrMetGluTyTyTyAlaAlaIleLysAsnAspGluPheMets	1242
Qy	7605	CCTTTGCAGAGATATGATGAAGCTGGAAACCATCATCTTCACGAAACTTAACACAAGAAC	7664
Db	1242	erPheValGlyThrTrpMetLysLeuGluThrIlelleLeuSerLysLeuSerGlnGluG	1262
Qy	7665	AGAAAAACCAACACCATGTTCTCACTTGTAAAGTGGGAGT	7705
Db	1262	lnLysThrLysHisArgIlePheSerLeuIleGlyLysAsn	1275

## RÉSUMÉ

GNRL1  
 retrovirus-related reverse transcriptase pseudogene - slow loris  
 C:Species: Nycticebus coucang (slow loris)  
 C:Date: 31-Mar-1998 #sequence\_revision 04-Jan-1996 #text\_change 14-May-1999  
 C:Accession: B25313  
 R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.  
 Nature 321, 625-628, 1986  
 A:Title: Ll family of repetitive DNA sequences in primates may be derived from a sequence  
 A:Reference number: A93381; MUID:86230917; PMID:2423883  
 A:Accession: B25313  
 A:Status: conceptual translation of pseudogene  
 A:Molecule type: DNA  
 A:Residues: 1-1260 <HAT>  
 A>Note: this sequence was constructed from an alignment of six sequences, determined  
 C:Keywords: reverse transcriptase; pseudogene

Alignment Scores:		
Pred. No.:	6.16e-282	Length: 1260
Score:	3658.50	Matches: 743
Percent Similarity:	66.90%	Conservative: 207
Best Local Similarity:	52.32%	Mismatches: 303
Query Match:	6.99%	Indels: 174
DB:	4	Gaps: 9

US-10-083-853B-2 (1-29921) x GNRL1 (1-1260)

[illegible]

QY	4154	GAACCTGAGAAATTAAAGAACTCACTCAAAATCACACAACTACATGGAACCTGAACAACCTG	4213	QY	5230	AGTCAAACTCCCTGAATATACACAGTAAACAGTTCTAAATTTGAAGCAGATTAATGTATAGCCT	5289
DB	235	GlueuAsnAsnAsnAsnAsnLeuHisThrHisThrLysThrLysThrLysLeuAsnAsnLeu	234	DB	444	uValGluMetLeuAsnArgProIleSerSerSerGluIleAlaSerThrIleGlnAsnLe	464
QY	4214	CTCCTGAATGACTACTGGGTAAATACAAAATGAAGCAAAAATAAAGATGTTCTTTGAA	4273	QY	5290	ACCAACCAAAAAAGTCCAGGACGAGCGATTTCACAGCCAAATTTTACCAGAGGTACAA	5349
DB	255	MetLeuLysAspThrTrpValIleAspGluIleLysLysGluIleThrLysPheLeuGlu	274	DB	464	uProLysLysLysSerProGlyProAspGlyPheThrSerGluPheThrGlnThrPheLys	484
QY	4274	ACCAATGAGACAAAGACAAATGTACCAAGAACTCTGGGCGCATATTTAAAGCAGTGTGT	4333	QY	5350	AGAGAAGCTGGTACTTCTCTGAACTATTCCAAAAATAGAAA--AATGGGAATCC	5406
DB	275	GlnAsnAsnGlnAspThrAsnTrpGlnAsnLeuTrpAspThrAlaLysAlaValLeu	294	DB	484	SGluGluLeuValProIleLeuLeuAsnLeuPheGlnAsnIleGluLysGlu-GlyIleL	504
QY	4334	AGAGGAAATTTATAGCACTAGATCGCTACAAGAGAAAGCAGGAAATATCTAAATAGAC	4393	QY	5407	TCCTTAACCTCATTTTACGAGCGCAGCATCATCTGTATACCAAAACCTAGCAGTGACACAA	5466
DB	295	ArgGlyLysPheIleAlaLeuGlnAlaPheLeuLysLysThrGluArgGluGluValAsn	314	DB	504	euProAsnThrPheTyrgluAlaAsnIleThrLeuIleProLysProGlyLysAspProt	524
QY	4394	ACCTTAACATCACAAATTAAGAAGCTAGAGAGAAAGCAACAATTCAAAAGCTAGC	4453	QY	5467	CAAAAAGAGGAATTTTCAGGCCCATATCCCTGTATGACATTTGATGTGAAAATCCTCAATA	5526
DB	315	AsnLeuMetGlyHisLeuLysGlnLeuGluLysGluHisSerAsnProLysProSer	334	DB	524	hrArgLysGluAsnTrpArgProIleSerLeuMetAsnIleAspAlaLysIleLeuAsnL	544
QY	4454	AGAAGACAGAAATAACTAAGATCAGAGCAAGTGAAGGAGATAGACACACAAAAGCC	4513	QY	5527	AAATACTTGGCAAAACCAATCCAGCAGCACATCAAAAAGCTTATCTACCATGATCAAGTTG	5586
DB	335	ArgArgLysGluIleThrLysIleArgAlaGluLeuAsnGluIleGluAsnLysArgIle	354	DB	544	ysIleLeuThrAsnArgIleGlnGlnHisIleLysLysIleIleHisAspGlnValG	564
QY	4514	CTTCAATTAATCAATCAATCCAGGAGCTGGTTTTTTGAAAAGATCAGCAAAATAGA--	4570	QY	5587	GCCTCATCCCTGGGATGCAAGGCTGGTTCAAAATATGCAAAATCAATAAATGTAGGCCATC	5646
DB	355	IleGlnGln-IleAsnLysSerLysSerTrpPheGluLysIleAsnLysIleAspLys	374	DB	564	lyPheIleProGlySerGlnGlyTrpPheAsnIleArgLysSerIleAsnValIleGlnH	584
QY	4571	-CCACTAGACAGACTAATAAAGAAGAAAGAGAGAGAAATCAAGAAGATGCAATAAAAA	4629	QY	5647	ACATAACAGAACCAATGACAAAACCAACATGATTTATCTCAATAGATGCGAGAAGGCCT	5706
DB	374	sProLeuAlaAsnLeuThrArgLysLysArgValLysSerLeuIleSerSerIleArgAs	394	DB	584	isIleAsnLysLysLysAsnLysAsnLysHisMetIleLeuSerIleAspAlaGluLysAlaP	604
QY	4630	TGATAAGGGGATATCACACCCATCCACAGACATCCACAGAAATACAACATATTATCAGAGAATATTA	4689	QY	5707	TTGTCAAAATTTCAACAGCCCTTCATGCTAAATAATTCACATAACTAGGTATCGATGAA	5766
DB	394	nGlyAsnAspGluIleThrAspProSerGluIleGlnLysIleLeuAsnGluTrpTy	414	DB	604	heAspAsnIleGlnHisPropHemIleArgThrLeuLysLysIleGlyIleGluGlyT	624
QY	4690	TAAACACCTCTATGCAATTAACATAGAAAATCTAGAGAATATGGATTAATCTCTGGACAC	4749	QY	5767	TGTATCTCAAAATTAATAAGAGCTATTATTATAC-AAACCCACAGCCAATATCATCTGAATG	5825
DB	414	rLysLysLeuTy-SerHisLysTyrgluAsnLeuLysGluIleAspGlnTyLeuGlu--	433	DB	624	hrPheLeuLysLeuIleGluAlaIleTySerLysProThrAlaAsnIleIleLeuAsnG	644
QY	4750	ATATGTAGCTGTATGGACCTTCGGGACAGACAAAGGGGTGAATGCAGAAATAAAA	4809	QY	5826	GGCAAAACCTGGAGCAATTCCTTTGAGAACTGGCACAACAGAGGATGCCCTCTCTCAC	5885
DB	433	-----	433	DB	644	lyValLysLeuLysSerPheProLeuArgSerGlyThrArgGlnGlyCysProLeuSerP	664
QY	4810	GACAAAGACAAAAGAGTATGTTTGAAGTAGGGGTAGGGGCAACTTGGCTCTAATGGA	4869	QY	5886	CACCTCTATTCAAGACTATTGTGAAGTTCTGGCCAGGCAATCAGCAAGATAGAGAAA	5945
DB	433	-----	433	DB	664	roleuLeuPheAsnIleValMetGluValLeuAlaIleAlaIleArgGluGluLysAlaI	684
QY	4870	CAAGGCCCTTGAGCTTTACACCACCTCTGTATTTATTAGGCAAAAGAGATAGCGAGG	4929	QY	5946	TAAAGGGTATTCAAAATAGAAAAGAGAGAGTATTTGTCTCTCTGTTTTCAGATGACATGT	6005
DB	433	-----	433	DB	684	leLysGlyIleHisIleGlySerGluGluIleLysLeuSerLeuPheAlaAspMetI	704
QY	4930	GTGAGTTGGAGAGAGGTGAGCTGTAGTCCAGAGTAGGCGCTGCAAGACTGCATTCCT	4989	QY	6006	TTGTATATTTAGAAAACCCCATCTCTCAGGCCAAAACCTCTTAACCTATAGCAACT	6065
DB	434	-----	436	DB	704	leValTyLeuGluAsnThrArgAspSerThrThrLysLeuLeuValIleLysGluT	724
QY	4990	CAAAACAATAGGCTCTAGATGTCCAGTAGATAACCTCAAGGAGCCAGTGCCAGGGAGTGA	5049	QY	6066	TCAGCAAAAGCTCAGGACACAAAATCAATGTGCAAAAATCACAAGCAATCTTATACGCCA	6125
DB	436	-----	436	DB	724	yrSerAsnValSerGlyTyrlLysIleAsnThrHisLysSerValAlaPheIleTyThrA	744
QY	5050	TGGCCCTCAGCAACCTTTAGGGCAGGCACAGAAGTAAGTTTGGCCACATCTGTATTTC	5109	QY	6126	ATAATAGACAAACAGAGAGCCAAATCATGATGAATCTCATTCACAAATTCGTACAAGA	6185
DB	436	-----	436	DB	744	snAsnAsnGlnAlaGluLysThrValLysAspSerIleProPheThrValValProLysL	764
QY	5110	ACGATAAACAGTTTGTGTTTGTATCAAGTAGCCCTCCAGTGGGAATGCTGAGTTGGTCATGA	5169	QY	6186	GAATAAAATACCTAGGATACACTTACAGGGCAGCGTAGGAACTCTTCAAGGAGAACT	6245
DB	436	-----	436	DB	764	ysMetLysTyrlLeuGlyValTyrlLeuThrLysAspValLysAspLeuTyrlLysGluAsnT	784
QY	5170	TCCTTTGGCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACACGAGA	5229	QY	6246	ACAAACCACTGATCAAGGAATAAGAGGACACACAAACAAATGGAACAACTCCATGCT	6305
DB	437	-----	444	DB	784	yrGluThrLeuArgLysGluIleAlaGluAspValAsnLysTrpLysAsnIleProcys	804
				QY	6306	CACAGATA-----GTAAGAATCATGAAAATGCG-ATACCTGCCCAAGTAATATATAGAT	6358

```
Db      804  erTrpLeuGlyArgIleAsnIleValLysMetSerIleLeuProLysAlaIleTyrAsnP 824
QY      6359 TCAGTGTACCCCAAGCTACCATTCACATTTCTTCCACAGAATTTGGAAAAACAACATT 6418
Db      824  heAsnAlaIleProIleLysAlaProLeuSerTyrPhelLysAspLeuGluLysIleIleL 844
QY      6419 TAAATTTTCATATGGAACCAAAAAAGAGCCACAGAGCCCAAGCAACATCTTAAGCAAAAAG 6478
Db      844  euHisPheIleIleTrpAsnGlnLysLys-ProGlnIleAlaLysThrLeuLeuSerAsnLys 863
QY      6479 AACAAAGCTGGAGGTATCATGCTACTGCTAACTAACTATATAAGCTACAGTAAC 6538
Db      864  AsnLysAlaGlyIleThrLeuProAspLeuArgLeuTyrTyrLysSerIleValIle 883
QY      6539 AAAAGCTGATGCTGCTGACCAAAACAGATATATAGCAACATGGAACAGACAGACC 6598
Db      884  LysThrAlaIleTyrTrpHisLysAsnArgLysValAspValTrpAsnArgIleGluAsn 903
QY      6599 TCAGAAATATAC-ACAGCAATCTACATCCATCTGATCTTGACAAACCTGACAAAAACAAG 6657
Db      904  GlnGluMetAspProAlaThrTyrHisTyrLeuIlePheAspLysProIleLysAsnIle 923
QY      6658 CAATGGAAAAAGATTCCTATTATTAATAATGGTGTGGAAAACTGGCTAGCCATATGC 6717
Db      924  GlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpValAsnTrpLeuAlaIleCys 943
QY      6718 AGAAAGCTGAAAGTCCCTCTCTACCTTATACAAAGTAACTCAAGATGAATT 6777
Db      944  ArgArgLeuLysLeuAspProHisLeuSerProLeuThrLysIleAspSerHisTrpIle 963
QY      6778 AAAGACTTAAATATAGACATAAAACCAATAAAACCCCA-GAAGAAACCTAGGCAATACC 6836
Db      964  LysAspLeuAsnLeuArgHisGlnThrIleLysIleLeuGluLysSerAlaGlyLysThr 983
QY      6837 ATTCAGGATATGGACATGGGCAACACTTCATGACTAAACCAACCAAGCAATGCCACA 6896
Db      984  LeuGluGlyIleSerLeuGlyLysTrpPheMetArgTrpGlnAlaIleGluAla 1003
QY      6897 AAAGCAAAATAGACAAGTGGGCTCTGATTAACATATAGAGCTCTCGCACAGCAAAAAA 6956
Db      1004 ValSerLysIleHisTyrTrpAspLeuLysLysLeuLysSerPheCysThrAla-LysAs 1023
QY      6957 AACTGTCTACGAGTGAACAAGCAACCTACAGAAATGGGAGAAATTTTGCATCTATCG 7016
Db      1023 nileValSerLysAlaSerArgGlnProSerGlnTrpGluLysIlePheAlaGlyTyrTh 1043
QY      7017 ATCTGACAAGCGTAAATATCCAGACATCTACGAACACTTAACAAATTTACAGAAAAA 7076
Db      1043 rSerAspLysGlyLeuIleThrArgIleHisArgGluLeuLysHisIleAsnLysLysAr 1063
QY      7077 AACAAAC---CCCGTCAAAATATGGCAAGGATATAGCAGACACTTCTCAAAAAGA 7132
Db      1063 gThrArgAspProIleSerGlyTrpAlaArgAspLeuLysArgAsnPheSerLysGluAs 1083
QY      7133 CATTTATCAGCAACAACATATGAAAAAACCTCATCATTCATTCGCTGTAGAGAAT 7192
Db      1083 pArgHisThrIleTyrLysHisMetLysLysSerSerSerLeuIleIleArgGluMe 1103
QY      7193 CCAAAACAACCAACGATGACATACCATCTCATCTGCTAGTATGATGGTGCATCAATAAA 7252
Db      1103 tGlnIleLysThrThrLeuArgTyrHisLeuThrProValArgValAlaHisIleThrLy 1123
QY      7253 GTCAGGAACAACAATGCTGGAGAGGATGTGGAGAAATAGAACACTTTTCCACTGTGTG 7312
Db      1123 sSerProAsnGlnArgCysTrpArgGlyCysGlyLysGlyThrLeuLeuHisCysTr 1143
QY      7313 GTGGGAATGTAATATAGTTCAACCACTTGTGGAAGACAGTGTGAGATTCCTTAAGGATCT 7372
Db      1143 pTrpGluCysProLeuIleArgSerPheTrpLysAspValTrpArgIleLeuArgAspLe 1163
QY      7373 AGAACCAAAATATCATTTGACCCAGCAATCCCATTTAGTATATACCAACCAAGGATA 7432
      1:::  ::::  |||||||  |||||||  |||||||  |||||||  |||  |
```

```
Db      1163 ulLysIleAspLeuPropheAspProIleIleProLeuLeuGlyLeuTyrProGluAspGl 1183
QY      7433 TAAATCATTTCTATTATAAGACACACATGCACATATGTTATTGACGACTGATCACAAT 7492
Db      1183 nLysSerGlnTyrAsnLysAspIleCysThrArgMetPheIleAlaIleGlnPheIleI 1203
QY      7493 AGCAAAAGACTTGGAAACCAACCCAAATGTCCATCAGTATAGATGGATGAAGAACAATG 7552
Db      1203 eAlaLysSerTrpLysLysProLysCysProSerThrHisGluTrpThrSerLysLeuTr 1223
QY      7553 GCACATATACACCATGAATACTATGACGCCATAAAAGGAT---GAGTTTCATGTCCTT 7608
Db      1223 pTyrMetTyrThrMetGluTyrTyrAlaAlaLeuLysLysAspGlyAspPheThrSerPh 1243
QY      7609 TCACAGATATGGAAGCTGGAACCAACCATCTCTCAGCAACACTTAACACAA 7660
Db      1243 eMetPheThrTrpMetGluLeuGluHisIleLeuLeuSerLysValSerGln 1260

RESULT 7
GNMSLL
retrovirus-related reverse transcriptase homolog - mouse retrotransposon
N:Alternate names: L1md repetitive element ORF-2; LINE-1 hypothetical protein; ORF 39
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 08-Jan-1999 #text_change 16-Jun-2000
C:Accession: B58927; B24906; I49130; A23772; B23430
R:Loeb, D.D.; Padgett, R.W.; Hardies, S.C.; Shehee, W.R.; Comer, M.B.; Edgell, M.H.;
Mol. Cell. Biol. 6, 168-182, 1986
A:Title: The sequence of a large L1md element reveals a tandemly repeated 5' end and
A:Reference number: A39072; MUID: 87064284; PMID: 3023821
A:Accession: B58927
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1281 <LOE1>
A:Cross-references: GB:M13002; NID:g200849; PIDN:AAA66024.1; PID:g804788
A:Note: sequence constructed using the first potential start codon for ORF2
A:Accession: B24906
A:Molecule type: DNA
A:Residues: 'NNQESHSTNOKEDSHKNR', 1-1281 <LOE2>
A:Cross-references: GB:M13002; NID:g200849
A:Note: sequence shown in Fig. 2
R:Martin, S.L.; Martin, S.L.
Gene 153, 261-266, 1995
A:Title: Characterization of a LINE-1 cDNA that originated from RNA present in ribonu
A:Reference number: I49129; MUID: 95180729; PMID: 7533116
A:Accession: I49130
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-85, 'L', 87-358, 'K', 360-706, 'R', 708-735, 'A', 737-760, 'W', 762-927, 'D', 929-1
A:Cross-references: EMBL:U15647; NID:g558906; PIDN:AAA67727.1; PID:g558908
R:Mottez, E.; Rogan, P.K.; Manueldis, L.
Nucleic Acids Res. 14, 3119-3136, 1986
A:Title: Conservation in the 5' region of the long interspersed mouse L1 repeat: impl
A:Reference number: A23772; MUID: 86176789; PMID: 3008107
A:Accession: A23772
A:Molecule type: DNA
A:Residues: 'NNQESHSTNOKEDSHKNR', 1-245, 'K', 247-423, 'SYTQONKWTWTWKN', 439, 'WTDTRYOS'
A:Cross-references: GB:X03725; NID:g52829; PIDN:CAA27363.1; PID:g1334115
C:Superfamily: pol polyprotein
C:Keywords: reverse transcriptase

Alignment Scores:
Pred. No.: 1.67e-273 Length: 1281
Score: 3552.50 Matches: 729
Percent Similarity: 65.39% Conservative: 208
Best Local Similarity: 50.87% Mismatches: 329
Query Match: 6.79% Indels: 174
DB: 1 Gaps: 9

US-10-083-853B-2 (1-29921) x GNMSLL (1-1281)
QY      3434 GGATCAATTCACACATAATAATTAACCTTAATGTAATGGCTAAATTCGCCAATTC 3493
Db      10  GlySerAsnAsnTyrPheSerLeuIleAsnGlyLeuAsnSerProIle 29
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```



```
QY 3494 AAAAGACACAGACTGGCAAAATTTGGATAAAGAGTCAAGACCCATCAGTGTGCTGATTTCAG 3553
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 30  LysArgHisArgLeuThrAspTfPLeuHisLysGlnAspProThrPheCysCysLeuGln 49
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3554 GAGGCCATCTCAGTCAAGAAAGACACACATAGGCTCAAAATAAAGGGATGGAGAGATT 3613
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 50  GluThrHisLeuArgGluLysAspArgHisTyrLeuArgValLysGlyTrpLysThrIle 69
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3614 TACCAAGTAATGAAACAAAAAAGAGGGGTTGCAATCTAGTCTCTGTATAA 3673
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 70  PheGlnAlaasnGly---LeuLysLysGlnAlaGlyValAlaIleLeuIleSerAspLys 88
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3674 ACAGACTTTAAACCAACAAAGATCAAAAGAGCAAAAGAGCCATTACATAATGGTAAAG 3733
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 89  IleAspPheGlnProLysValIleLysLysAspLysGluGlyHisPheIleLeuIleLys 108
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3734 GCATCAATGGAACAGAGCTAACTATCTTAATATACATGCACCCCAATACAGGACA 3793
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 109 GlyLysIleLeuGlnGluLeuSerIleLeuAsnIleTyrAlaProAsnAlaArgAla 128
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3794 CCCAGATTCAAAAGCAAGTTCTTAGAGACCTCAAAAGACATTTGACTCCACACAAATA 3853
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 AlaThrPheIleArgAspThrLeuValLysLeuLysAlaTyrIleAlaProHisThrIle 148
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3854 ATAGTGGAGTCTAAATAATAATAGACACTTTAAACACCCCACTGCCAATATTAGGCAGA 3913
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 149 IleValGly-----AspPheAsnThrProLeuSerSerLysAspArg 162
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3914 TCAATGACAGACAAAATTACAGGATATCCAGAGTTGAAGTTCAGCTCTGGACCAAGCC 3973
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 163 SerTrpLysGlnLysLeuAsnArgAspThrValLysLeuThrGluValMetLysGlnMet 182
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3974 GACTTAATAGATATCAGAACTCCCAACCCCAATCAACAGAAATACACATCTCTCTCA 4033
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 183 AspLeuThrAspIleArgThrPheTyrProLysThrLysGlyTyrThrPhePheSer 202
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4034 GCATCAATACACTATTTTAAATTTGACCATGTAAATTTAAAGTAAACACCTCCTCAGC 4093
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 203 AlaProHisGlyThrPheSerLysIleAspHisIleIleGlyHisLysThrGlyLeuAsn 222
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4094 AAATGCAAGACAGACAAATCTCAACAAACAGTCTCTCAGACACAGTGAATCTATTTA 4153
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 223 ArgTyrLysAsnIleGluIleValProCysIleLeuSerAspHisLysGlyLeuArgLeu 242
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4154 GAAGTCAGAAATTAAGAACTCACTCAAAATCACACAACTACATGGAACTGAACAACCTG 4213
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 243 IlePheAsnAsnAsnIleAsnAsnGlyLysProThrPheThrTrpLysLeuAsnAsnThr 262
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4214 CTCTGAATGACTGGTAAATAACAAAATGAAGGCAAAAATAAAGATGTTCTTTGAA 4273
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 263 LeuLeuAsnAspThrLeuValLysGluGlyIleLysLysGluIleLysAspPheLeuGlu 282
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4274 ACCAATGAGAACAAAGACAAATGTACCAGAACTCTGGGGCATATTTAAAGCAGTGTG 4333
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 283 PheAsnGluAsnGlnAlaThrThrTyrProAsnLeuTrpAspThrMetLysAlaPheLeu 302
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4334 AGAGGGAATTTATAGCACTAGATGCTTACAGAGAAAGCAGGAATATCTAAATAGAC 4393
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 303 ArgGlyLysLeuIleAlaLeuSerAlaSerLysLysArgGluThrAlaHisThrSer 322
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4394 ACCTTACATCACAATTAAGAACTAGAGAAGAAAGCAACAAATTCAAAAGCTAGC 4453
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 323 SerLeuThrThrHisLeuLysAlaLeuGluLysLysGluAlaAsnSerProLysArgSer 342
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4454 AGAAGACAAAGAAATACTAGATCAGACGAACTGAAGGATAGATAGACACAAAAGCC 4513
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 343 ArgArgGlnGlnIleLysLeuArgGlyGluIleAsnGlnValGluThrArgArgThr 362
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4514 CTTCAATAATCAATCAATCCAGGAGCTGTTTTTGAAGAAATCAGCAAAATAGA--- 4570
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 363 IleGlnArg-IleAsnGlnThrArgSerTrpPheGluLysIleAsnLysIleAspLys 382
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
Db 592 yrlAsnLysLeuLysAspLysAsnHisMetIleLeSerLeuAspAlaGluLysAlap 612
:::|||||::: |||||||
QY 5707 TTGTCRAAATCAACAGCCCTTCATGCTAAATATCTCAGTAACATAGGTATGATGGAA 5766
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 612 heAspIleGlnHisProPheMetIleLysValLeuGluArgSerGlyIleGlnGlyP 632
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 5767 TGTATCTCAAAATAAATAAGAGCTATTATATAC-AAACCCACAGCCAATATCATCTGAATG 5825
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 632 rofYrLeuAsnMetIleLysAlaIleYrSerLysProValAlaAsnIleLysValAsnG 652
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 5826 GGCAGAAATCGAAGCATTCCTTTGAGAACTGGCACAGCAAGGATGCCCTCTCTCAC 5885
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 652 lyGluLysLeuGluAlaIleProLeuLysSerGlyThrArgGlnGlyCysProLeuSerP 672
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 5886 CACTCTATTCAAGATACATTGGAAGTCTGCGCCAGGCAATCAGGCAATAGAAAGAA 5945
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 672 rofYrLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGlnLysGluI 692
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 5946 TAAAGGTTATTCAAAATAGAAAGAGAGAGTATGCTCTGTTGTCAGATGACATGT 6005
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 692 leLysGlyIleGlnIleGlyLysGluGluValLysIleSerLeuLeuAlaAspMetI 712
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6006 TTGTAATATTAGAAAACCCATCGCTCTCAGGCCAAAACCTCTTAAGCTGATAGCAACT 6065
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 712 leValYrIleSerAspProLysAsnSerThrArgGluLeuLeuAsnLeuIleAsnSerP 732
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6066 TCAGCAAACTCAGGACACAAATCAATGCTGCAAAATCACAAGCATCTCTATAGCCCA 6125
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 732 heGlyIleValValGlyYrLysIleAsnSerAsnLysSerMetAlaPheLeuYrThrL 752
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6126 ATAATAGACAAACAGAGAGCCAAATCATGAGTGAACCTCTCATTCACAAATGTCTACAAGA 6185
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 752 ysAsnLysGlnAlaGluLysGluIleArgGluThrThrProPheSerIleValThrAsnA 772
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6186 GAATAAATACCTAGGAATACAACTTACAGGGACAGCTAGGAACCTCTCAGGAGAACT 6245
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 772 snIleLysYrLeuGlyValThrLeuThrLysGluValLysAspLeuYrAspLysAsnP 792
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6246 ACAAAACCATGATCAAGAAATAAAGAGAGACACAAACAAATCGGAAAAACATTCATGCT 6305
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 792 heLysSerLeuLysLysGluIleLysGluAspLeuArgYrGlyLysAspLysProCysS 812
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6306 CACAGATAGTAAGAATCAT-----GAAAATGCCACTGCCCCAAAGTAAATATATAGAT 6358
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 812 erTriPLeGlyArgIleAsnIleValLysMetAlaIleLeuProLysAlaIleYrArgP 832
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6359 TCAGTGCTACCCCATCAAGCTACCATTCATCTTCTTCAGAAATTTGGAACAAACACTT 6418
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 832 heAsnAlaIleProIleLysIleProThrGlnPheAsnGluLeuGluGlyAlaIleC 852
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6419 TAAATTTTCATATGGAACCAAAAAGAGAGCCACAGAGCCAAAGACAATCTTAAAGCAAAAAG 6478
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 852 ysLysPheValTrpAsnAsnLysLys-ProArgIleAlaLysSerLeuLeuLysAspLys 871
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6479 ACAAGCTGGAGGTATACGTACTGCTGACTTAAACTATACTATTAAGCTACAGTAACC 6538
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 872 ArgThrSerGlyIleThrMetProAspLeuLysLeuYrYrArgAlaIleValIle 891
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6539 AAACTGCGATGGTACTGGTACCAAAACAGATATATAGCAATGGAACACAGAGACC 6598
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 892 LysThrAlaTrpYrTrpYrArgAspArgGlnValAspGlnTrpAsnArgIleGluAsp 911
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6599 TCAGAAATACACT-GCAATCTACATCCATCTGCTTTGACAAACCTGACAAAAACAAG 6657
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 912 ProGluMetAsnProHisThrYrGlyHisLeuIlePheAspLysGlyAlaLysThrIle 931
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6658 CAATGAAAAGGATTCCTATTAATAATGGTGTGGAAAACCTGGCTACGCATATGC 6717
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 932 GlnTrpLysLysAspSerIlePheAsnAsnTrpCysTrpHisAsnTrpLeuLeuSerCys 951
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6718 AGAACTGAACTGATGCTCTCTTACACTTATACAAAGATTAACATCAAGATGAAT 6777
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

## RESULT 8

S23650

retrovirus-related hypothetical protein II - human retrotransposon LINE-1

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence\_revision 01-Nov-1996 #text\_change 08-Jan-1999

C:Accession: S23650

R:HohJoh, H.; Minakami, R.; Sakaki, Y.

```
Db 952 ArgArgMetArgIleAspProYrLeuSerProCysThrLysValLysSerLysTrpIle 971
QY 6778 AAAGACTTAAATATAGACATAAACCAATAAAACCCCA-GAAGAAACCTAGCAATACC 6836
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 972 LysGluLeuHisIleLysProGluThrLeuLysLeuIleGluLysValGlyLysSer 991
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6837 ATTCAAGATATGACATGGCCAAAGACTTCATCAGCTAAAAACCAACCAATGGCAACA 6896
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 992 LeuGluAspMetGlyThrGlyGluLysPheLeuAsnArgThrAlaMetAlaCysAlaVal 1011
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6897 AAAGCCAAATAGACAAGTGGGATCTGATTAACATATAGAGCTTCGACAGCAAAAAA 6956
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1012 ArgSerArgIleAspLysTrpAspLeuMetLysLeuGlnSerPheCysLysAla-Lysas 1031
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6957 AACTGTCATCAGAGTCAACAAGCAACCTACAGAAATGGAGAAATTTTGCATCTATCG 7016
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1031 pthrValAsnLysThrLysArgProProThrAspTrpGluArgIlePheThrYrProLy 1051
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7017 ATCTGACAAAGGCTAATATCCAGAGATCTACGAAGAACTTAAACAAATTTA----CAAGA 7072
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1051 sSerAspArgGlyLeuIleSerAsnIleYrLysGluLeuLysLysValAspPheArgLy 1071
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7073 AAAAAACAACCCGCTCAAAATATGGCAAGGATATGACAGACACACTTCTCAAAACAAGA 7132
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1071 sSerAsnAsnProIleLysLysTrpGlySerGluLeuAsnLysGluPheSerProGluI 1091
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7133 CATTTATGAGCAACAACAATATGAAAAACCTCATCATCTGCTGCTAGAGAAAT 7192
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1091 uYrArgMetAlaGluLysHisLeuLysLysCysSerThrSerLeuIleIleArgGluMe 1111
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7193 GCAAAACAACACCAAGTGCATCATCTCATCTAGTGTAGTAATGGTGATCACTAAAAA 7252
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1111 tGlnIleLysThrLeuArgPheHisLeuThrProValArgMetAlaLysIleLysas 1131
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7253 GTCAGAAACAACAATATGCTGGAGAGGATGTGGAGAAATAGGNACACTTTCCACTGTTC 7312
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1131 nSerGlyAspSerArgCysTrpArgGlyCysGlyGluArgGlyThrLeuLeuHisCysTr 1151
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7313 GTGGGAATGTAATAGTTCAACCATTTGGAGACAGTGTGGAGATTCCTTAAGGATCT 7372
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1151 pTrpGluCysArgLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArgLysLe 1171
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7373 AGAACAGAAATATATATTTGACCCCAATCCATTCAGTATATATATATATATATATAT 7432
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1171 uAspIleValLeuProGluAspProAlaIleProLeuLeuGlyIleYrProGluAspAl 1191
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7433 TAAATCATCTATTATATAAGACACATGACACATATGTTTATTTGACAGCACTGATCACA 7492
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1191 aProThr----GlyLysLysAspThrCysSerThrMetPheIleAlaLeuPheIleI 1210
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7493 AGCAAGACTTGGAAACCAACCAATGTCCATCAGTATAGACTGGATATAAGAAAAACATG 7552
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1210 eAlaArgSerTrpLysGluProArgCysProSerThrGluGluTrpIleGlnLysMetTr 1230
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7553 GCACATATACACCATGAATATCTATGACGCCAT-AAAAAGATGAGTTCATGTCCTTTCG 7611
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1230 pYrIleYrThrMetGluYrYrSerAlaIleLysLysAsnGluPheMetLysPheLe 1250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7612 AGAGATATGATGAGCTGGAACCATCTCTCAGCAAACTAACAACAAGACAGAAAAAC 7671
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1250 uAlaLysTrpMetAspLeuGluGlyIleIleLeuSerGluValThrHisSerGlnArgas 1270
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 7672 CAAAACCAACATGTTCTCTACATCTGTAAGTGGG 7702
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1270 nSerHisAsnMetYrSerLeuIleSerGly 1280
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

Nucleic Acids Res. 18, 4099-4104, 1990  
 A:Title: Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which  
 A:Reference number: S23649; MUID:90332398; PMID:2165587

A:Accession: S23650  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A:Residues: 1-712 <OH>

A:Cross-references: EMBL:X52235

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1990

C:Genetics:

A:Mobile element: LINE-1

A:Start codon: GTG

C:Superfamily: pol polyprotein

Alignment Scores:  
 Pred. No.: 3,85e-220 Length: 712  
 Score: 2880.50 Matches: 573  
 Percent Similarity: 87.11% Conservative: 49  
 Best Local Similarity: 80.25% Mismatches: 87  
 Query Match: 5.50% Indels: 11  
 DB: 2 Gaps: 2

US-10-083-853B-2 (1-29921) x S23650 (1-712)

Qy 5583 GTTGGCTCATCCCTGGGATGCAAGGCTGTTCAAAATATGCAAAATCAATAATGTAGGC 5642  
 Db 1 MetGlyPheLeuProGlyMetGlnAspTriPheAsnMetHisLysSerIleAsnValIle 20  
 Qy 5643 CATCACATAACAGAACCAATGACAAAACACATGATTATCTCAATAGATGCAAGAAAG 5702  
 Db 21 GlnHisIleAsnArgThrLysAspLysAsnHisMetIleValSerIleAspAlaGluLys 40  
 Qy 5703 GCCTTTCTCAAAATTAACAGCCCTTCATGCTGTAATAATCTCAGTAACACTAGGTATCGAT 5762  
 Db 41 AlaPheAspLysIleGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleAsp 60  
 Qy 5763 GGAATGATATCTCAAAATTAAGAGCTATTATATAC - AAACCCACAGACCAATATCATCTG 5821  
 Db 61 GlyThrTyrrPheLysIleIleArgAlaIleTyrrAspLysProThrAlaAsnIleIleLeu 80  
 Qy 5822 AATGGCAAAACCTGGAGCATTCCTTGGAGACTGGCACAAGCAAGGATGCCCTCTC 5881  
 Db 81 AsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProLeu 100  
 Qy 5882 TCACCATCTCTATTCAAGATACTATTGGAAGTCTGGCCAGGGAATCAGGCAATAGAA 5941  
 Db 101 SerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLys 120  
 Qy 5942 GAAATAAGGATATCAAAATAGAAAGAGAGAGTATATGCTCTCTTTGCGAGATCAC 6001  
 Db 121 GluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspAsp 140  
 Qy 6002 ATGTTTCTATATTAGAAACCCCATCTCTCAGGCCCAAAACCTCTTAAGCTATAGC 6061  
 Db 141 MetIleValTyrrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysProIleSer 160  
 Qy 6062 AACTTCAGCAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAAGCATCTTATAC 6121  
 Db 161 AsnPheSerLysValSerGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeuTyrr 180  
 Qy 6122 GCCAATATAGCAAAACAGAGACCAATCATGATGATGAATCTCATTCATTCACAAATTCGAC 6181  
 Db 181 ThrAsnAsnArgGlnThrGluSerGlnIleMetAsnGluLeuProPheThrIleAlaSer 200  
 Qy 6182 AAGAAATAAATACCTAGGAATACAACTTACAGGACACGCTAGGAACTCTTCAAGGAG 6241  
 Db 201 LysArgIleLysTyrrLeuGlyIleGlnIleThrArgAspValLysAspLeuPheLysGlu 220  
 Qy 6242 AACTTACAAACACCTGATCAAGGAATAAGAGGACACAAACAAATGGAAAAACATTCCA 6301  
 Db 221 AsnTyrrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTyrrLysAsnIlePro 240  
 Qy 6302 TGCTCACAGATA-----GTAAGAATCATGAAATGCC-ATACTGCCCAAGTAATATAT 6354

Db 241 CysSerTrpValGlyArgIleAsnIleMetLysMetAlaIleLeuProLysValIleTyrr 260  
 Qy 6355 AGATTTCAGTCTACCCCATCAAGCTACCATCTTCTTCACAGATTCGGAAGAAACA 6414  
 Db 261 ArgPheAsnThrIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysThr 280  
 Qy 6415 ACTTTAAATTTTCATATGGAACCAAAAGAGCCACAGAGCCCAAGCAATCTTAAAGCAA 6474  
 Db 281 ThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysAlaIleArgSerG1 300  
 Qy 6475 AAAGAACAAAGCTGGAGGTATCATCTACCTGACTTAAACTATACTATAGGTCTACAGT 6534  
 Db 300 nLysAsnLysSerGlyGlyIleThrLeuProAspPheLysLeuTyrrTyrrLysAlaThrVa 320  
 Qy 6535 AACCAAAACTGCATGGTACTGGTACCAAAACAGATATATACCAAAATGGAACACAGACA 6594  
 Db 320 lThrLysThrAlaTrpTyrrTyrrGlnAsnArgAspIleAspGlnTrpAsnArgThrG1 340  
 Qy 6595 GACCTCAGAAATTACACT-GCAATCTACATCCATCTGATCTTTCACAAACCTGACAAAAA 6653  
 Db 340 uProSerGluIleThrProHisIleTyrrAsnTyrrLeuIlePheAspLysProGluLysAs 360  
 Qy 6654 CAAGCAATGGAAGAGATCCCTATTATTAATAATGGTGTGGAAAACTGGCTAGCCAT 6713  
 Db 360 nGluGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaI1 380  
 Qy 6714 ATGCAGAAAGCTGAAGCTGGATCCCTTCTTACACCTTATATACAAAGTTACTCAAGATG 6773  
 Db 380 eCysArgLysLeuLysLeuAspProPheLeuThrProTyrrThrLysIleAsnSerArgTr 400  
 Qy 6774 AATTAAGACTTAATATAGACATAAAACCAATAAAACCCCA-GAAGAAAACTTAGGCAA 6832  
 Db 400 PileLysAspLeuIleValArgProLysThrIleLysThrLeuGluGluAsnLeuGlyI1 420  
 Qy 6833 TACCATTGAGGATATGACATGGCAAGACTTCATGACTATAACACACAAAGCAATGGC 6892  
 Db 420 eThrIleGlnAspIleGlyMetGlyLysAspPheMetSerLysThrProLysAlaMetal 440  
 Qy 6893 AACAAAAACCAAAATAGACAGTGGATCTGATTAAACTATATAGCTTCTCTCAGACGAAA 6952  
 Db 440 aThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla-L 460  
 Qy 6953 AAAAACTGTCTATCAGAGTGAACAGCAACCTACAGATGGGAGAAAATTTTTCGAATCT 7012  
 Db 460 ySGLuThrThrIleArgValAsnArgGlnProThrLysTrpGluLysThrPheAlaThr 480  
 Qy 7013 ATCGATCTGCAAAAGGCTAATATCCAGAGATCTACGAGAACTTAACAAATTTTACAAGA 7072  
 Db 480 yrSerSerAspLysGlyLeuIleSerArgIleTyrrAsnGluLysGlnIleTyrrLysL 500  
 Qy 7073 AAAA---AACAAACCCGTCAAAATATGGGCAAGGATATGAGACAGACACTTCTCAAAAG 7128  
 Db 500 yLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLysG 520  
 Qy 7129 AAGACATTTATGACGCCCAACAAACATATGAAAAAACCTCATCATCTATGCTGTAGAG 7188  
 Db 520 luAspIleTyrrAlaAlaLysLysHisMetLysLysCysSerProSerLeuAlaIleArgG 540  
 Qy 7189 AAATGCAAAACAAACACAGTGCATACCATCTCATCTAGTGTAGTGTAGTGTAGTACTA 7248  
 Db 540 luMetGlnIleLysThrThrMetArgTyrrHisLeuThrProValArgMetAlaIleLe 560  
 Qy 7249 AAAAGTCAGGAAACAAACAAATGCTGGAGAGATCTGGGAATAGGACACACTTTTCCACT 7308  
 Db 560 yLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGlyIleGlyIleGlyThrLeuLeuHsc 580  
 Qy 7309 GTTGGTGGGATGTAAATTAGTTCACATTGTGGAAGACAGAGTGTGGAGATTCCTTAAGG 7368  
 Db 580 ySTrptAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuArg 600  
 Qy 7369 ATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCCATCTACTGATGATATACCAAGG 7428  
 Db 7428 ATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCCATCTACTGATGATATACCAAGG 7428

600 spLeuGluProGluIleProPheAspProAlaIleProLeuLeuGlyIleTyrProLysA 620

7429 AATAATAAATCATTTATTATAAAGACACATGACACATATATGTTATTGCGAGCTATGATCA 7488

620 spSerLysSerCysTyrLysAspThrCysThrArgMetPheIleAlaLeuPheT 640

7489 CAATAGCAAAAGACTTGGAAACCAACCCAAATGTCCATCAGTGATAGACTGGATAAAGAAA 7548

640 hrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLysM 660

7549 CATGGGCACATATACACCATGAATACTATGCGACCCATAAAAG-GATGATGTTTCATGTCCT 7607

660 etrPheHisIleTyrThrMetGluTyrAlaIleLysAsnGluPheValserP 680

7608 TTGCAGAGATATGGATGAAGCTGGAACCATCATCTTCAGCAAACTAAACACAAAGACAGA 7667

680 heValGlyThrTrpMetLysLeuGluIlelleLeuSerLysLeuSerGlnGluGlnL 700

7668 AAACCAACACCATCTTCTCACTTGTAAAGTGGGAGT 7705

700 ysthrLysHisCysIlePheSerLeuIleGlyGlyAsn 712

RESULT 9

B34087

hypothetical protein (L1H 3' region) - human

C:Species: Homo sapiens (man)

C:Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 30-Sep-1993

C:Accession: B34087

R:Scott, A.F.; Schmeckpeper, B.J.; Abdelrazik, M.; Comey, C.T.; O'Hara, B.; Rossiter, J. J. Genomics 1, 113-125, 1987

A:Title: Origin of the human L1 elements: proposed progenitor genes deduced from a conserved

A:Reference number: A34087; MUID:88085185; PMID:3692483

A:Accession: B34087

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1280 <SC0>

C:Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1-79e-170	Length:	1280
Score:	2257.00	Matches:	412
Percent Similarity:	96.15%	Conservative:	12
Best Local Similarity:	93.42%	Mismatches:	17
Query Match:	4.31%	Indels:	0
DB:	2	Gaps:	0

US-10-083-853B-2 (1-29921) x B34087 (1-1280)

Qy 29919 ACTTCTTCAGAAATTTGGAAAACTACTTTAAAGTTCACTGTGGAACCAAAAAAGAGCC 29860

Db 840 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleThrPasnGlnLysArgAla 859

Qy 29859 CAGATGCCCAAGTCAATCCTTAAGCCAAAGACAAAGCTGGAGGCATCATGCTACCTGCAC 29800

Db 860 HisIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAsp 879

Qy 29799 TTCAAACTACACTACAGGCTACAGTAACCAACCAACATGGTACTGGTACCAAAACAGA 29740

Db 880 PheLysLeuTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 899

Qy 29739 GATATCAACCAATGGAAACAGACAGCCCTCAGAATAATGCCGATATCTACAACATAT 29680

Db 900 AspileAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrAsnTyr 919

Qy 29679 CTGATCTTTGACAAACCTGAGAAAAACAAGCAATGGGAAAGGATTCCTCTATTATAATAA 29620

Db 920 LeuIlePheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 939

Qy 29619 TGCTGCTGAGAAACCTGGCTAGCCCTATCTAGAAAGCTGAACTGATCCCTTCCCTTACA 29560

Db 940 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 959

Qy 29559 CCTTATACAAAAATTAATTCAGATGATTAAAGACTTAAATGTTAGACCTAAACACCAATA 29500

Db	960	ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle	979
QY	29499	AAATCCCTAGAGAAACCCAGGCAATACACATTACAGGACATAGGCATGGCAAGGACTTC	29440
Db	980	LysThrLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe	999
QY	29439	ATGTCTAAAACACCAAAAGCAATGGCAACAAAAGCCAAAATTCACAAATGGGATCTAAAT	29380
Db	1000	MetThrLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle	1019
QY	29379	AAACTAAGAGCTTCGCACAGCAAAAGAAACTACATCAGAGTGAAACAGGCAACTTACA	29320
Db	1020	LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr	1039
QY	29319	GAATGGGAGAAAATTTTGGCAACCTACTCATTTGACAAAGGCGTAAATATCCAGAATCTAC	29260
Db	1040	GluTrpGluLysIlePheAlaIleTyrSerSerAspLysGlyLeuIleSerArgIleTyr	1059
QY	29259	AATGAACCTCAAAACAATTTTACAAGAAAAAACAAGAACCCCAATCAAAAAGTGGGTGAAG	29200
Db	1060	AsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLys	1079
QY	29199	GATATGAACAGACACTTCTCAAAAGAGGCGATTTATGCAGGCCCAAAAAACACATGAAAAA	29140
Db	1080	AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaAsnLysHisMetLysLys	1099
QY	29139	TACTCATCATCACTGGCCATCACAGAAATGCAAAATCAAAACACCAATGAGATACCATCTC	29080
Db	1100	CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu	1119
QY	29079	ACACGATTAGAATGGCAATCATTAATAAAGTCAGGAAACAACAGGTGCTGGAGAGAATGT	29020
Db	1120	ThrProValargMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys	1139
QY	29019	GGAGAAATAGAACACTTTTACACTGTTGGTGGGACTGTGAAGTCTCAACCATCTGTGG	28960
Db	1140	GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp	1159
QY	28959	AAGTCAGTGGTGATTCTCTCAGGGATCTAGAACTAGAAATACCATTTTCAACCCAGCCATC	28900
Db	1160	LysThrValTrpArgPheLeuLysAspLeuGluLeuGluIleProPheAspProAlaIle	1179
QY	28899	CCATTACTGGGTATATACCCAAAGGATTATAATCATGTGCTGTATAAAGACACAAGACACA	28840
Db	1180	ProLeuLeuGlyIleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThr	1199
QY	28839	TGTATGTTTATAGCAGCACTATTACAATAGCAAAAGACTTGGAAACCAACTAAATGTCCA	28780
Db	1200	ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro	1219
QY	28779	ACAAGATACACTGGATTAAAGAAATGTGGCACAATATACACCATGGAATACTATCGAGCC	28720
Db	1220	SerMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTrpAlaAla	1239
QY	28719	ATAAAAATCATGATTGCTGCTCTTTGTAGGGACATGGATGAAGCTGGAACCACTCAAT	28660
Db	1240	IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIleIle	1259
QY	28659	CTCAGCAAACTATCACAAAGSACAAAAACCAACCCGATGTTCTCACTCATAGTGGG	28600
Db	1260	LeuSerLysLeuSerGlnGlyGlnLysThrLysHisArgMetPheSerLeuIleGlyGly	1279
QY	28599	AAT 28597	
Db	1280	Asn 1280	
RESULT 10			
S65824			
reverse transcriptase homolog - human transposon L1.1			
C:Species: Homo sapiens (man)			
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999			
C:Accession: S65824			

R:Dombroski, B.A.  
submitted to the EMBL Data Library, January 1992  
A:Description: Isolation of an active human transposable element.  
A:Reference number: S65823

A:Accession: S65824

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1275 <D>

C:Cross-references: EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PID:g339771

C:Superfamily: pol polyprotein

#### Alignment Scores:

Pred. No.:	3,01e-168	Length:	1275
Score:	2229.00	Matches:	409
Percent Similarity:	95.24%	Conservative:	11
Best Local Similarity:	92.74%	Mismatches:	21
Query Match:	4.25%	Indels:	0
DB:	2	Gaps:	0

US-10-083-853B-2 (1-29921) x S65824 (1-1275)

QY	29919	ACTTTCTTCCACAGAAATGCAAAACTACTTTAAAGTTTCATGTGGAAACCAAAAAAGAGCC	29860
Db	835	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla	854
QY	29859	CACATCGCCCAAGTCAATCCTAAGCCAAAGAAACAAAGCTGGAGGCATCATGCTACCTGAC	29800
Db	855	ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp	874
QY	29799	TTCAAACTACACTACAGCTCAGTAACCAAAACAACTAGTACTGGTACCAAAACAGA	29740
Db	875	PheLysLeuTyrrLysAlaThrValThrLysThrAlaTrpTyrrTrpTyrrGlnAsnArg	894
QY	29739	GATATCAACCAATGGAACAGACAGAGCCCTCAGAAATATCCGCATATCTACACTAT	29680
Db	895	AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrrAsnTyrr	914
QY	29679	CTGATCTTTGACAAACCTGAGAAACAAAGCAATGGGAAAGATTCCTATTTAATAAA	29620
Db	915	LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	934
QY	29619	TGGTGCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAACTGGATCCCTTCCTTACA	29560
Db	935	TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr	954
QY	29559	CCTTATACAAATTAATTCAGATGGATTAAAGACTTAAATCTTAGACTTAAACCAATA	29500
Db	955	ProTyrrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValLysProLysThrIle	974
QY	29499	AAATCCTTGAAGAAACCCAGGCAATACCATTCAGGACATAGGCATGGCAAGGACTTC	29440
Db	975	LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe	994
QY	29439	ATGCTTAAACACCAAAACCAATGGCAACAAAGCCAAATTCACAAATGGGATCAAT	29380
Db	995	MetSerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuIle	1014
QY	29379	AACTAAAGAGCTTCTGCACAGCAAGAAACTACCATCAGAGTGAACAGGCAACTTACA	29320
Db	1015	LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr	1034
QY	29319	GAATGGGAGAAATTTTTCACACCTACTCATTTGACAAAGGGCTAATATCCAGAACTAC	29260
Db	1035	ThrTrpGluLysIlePheAlaThrTyrrSerAspLysGlyLeuIleSerArgIleTyrr	1054
QY	29259	ATGAACTCAACAAATTTACAGAAACAAACAAAGAACCCCATCAAAAGTGGGTGAAG	29200
Db	1055	AsnGluLeuLysGlnIleTyrrLysLysLysThrAsnAsnProIleLysLysTrpAlaLys	1074
QY	29199	GATATGACAGACACTTCTCAAGAGAGGCATTTATGCAGCCCAAAAAACACATGAAAA	29140
Db	1075	AspMetAsnArgHisPheSerLysGluAspIleTyrrAlaAlaLysLysHisMetLysLys	1094

QY	29139	TACTCATCATCACTGGCCCATCAGAGAAATGCAAAATCAAAACCAACCAATGAGATACCATCTC	29080
Db	1095	CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrrHisLeu	1114
QY	29079	ACACCAAGTTAGATGGCAATCATTTAAAAAGTTCAGGAAACAAACAGGTGCTGGAGAGAAATGT	29020
Db	1115	ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys	1134
QY	29019	GGAGAAATAGGACACTTTTACACTGTTGTTGGGACTGTCAACTACTTCAACCATTTGGTGG	28960
Db	1135	GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp	1154
QY	28959	AACTCAGTGTGGTATTCTCCTCAGGATCTAGAACTAGAAATACCAATACCTTTGACCCAGCCATC	28900
Db	1155	LysSerValTrpArgPheLeuArgAspLeuGluLeuIleProPheAspProAlaIle	1174
QY	28899	CCATTACTGGGTATATACCCAAAGGATTAATAATCATGTGCTGTATTAAGACACACACACACA	28840
Db	1175	ProLeuLeuGlyIleTyrrProAsnGluTyrrLysSerCysCysTyrrLysAspThrCysThr	1194
QY	28839	TGTATGTTTATAGCAGCACTATTCACATAGCAAAAGACTTGGAAACCACTTAATCTCCA	28780
Db	1195	ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro	1214
QY	28779	ACAACCATAGACTGGATTAAAGAAATGTGCACATATACACCATGGAATCTATGACGCC	28720
Db	1215	ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrrThrMetGluTyrrAlaAla	1234
QY	28719	ATAAAAATGATGAGTTCCTGCTCTTTAGGACATGGATGAAGCTGGAACCATCATTT	28660
Db	1235	IleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrIleIle	1254
QY	28659	CTCAGCAAACTATCACAGGACAAACCAACACCGCATGTCTCTCACTCATAGTGGG	28600
Db	1255	LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGlyGly	1274
QY	28599	AAT 28597	
Db	1275	Asn 1275	
RESULT 11			
I38588			
reverse transcriptase homolog - human retrotransposon L1			
N:Alternate names: ORF2 protein			
C:Species: Homo sapiens (man)			
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 26-Aug-1999			
C:Accession: I38588			
R:Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.			
A:Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1			
A:Reference number: I38587; MUID:95004577; PMID:7920631			
A:Accession: I38588			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1275 <RES>			
A:Cross-references: EMBL:U09116; NID:g483914; PIDN:AA60345.1; PID:g483916			
C:Superfamily: pol polyprotein			
Alignment Scores:			
Pred. No.:	3,91e-167	Length:	1275
Score:	2215.00	Matches:	408
Percent Similarity:	95.01%	Conservative:	11
Best Local Similarity:	92.52%	Mismatches:	22
Query Match:	4.23%	Indels:	0
DB:	2	Gaps:	0
US-10-083-853B-2 (1-29921) x I38588 (1-1275)			
QY	29919	ACTTTCTTCCACAGAAATGCAAAACTACTTTAAAGTTTCATGTGGAAACCAAAAAAGAGCC	29860
Db	835	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla	854
QY	29859	CACATCGCCCAAGTCAATCCTAAGCCAAAGAAACAAAGCTGGAGGCATCATGCTACCTGAC	29800

Db	855	ArglleAlaLysSerlleLeuSerGlnLysAsnLysAlaGlylleThrLeuProTyr	874
Qy	29799	TTCAAACTACACTCAAGGCTACAGTAACCAAAACCAACATGGTACTGGTACCAAAACAGA	29740
Db	875	PhelysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTyrTyrGlnAsnArg	894
Qy	29739	GATATCAACCAATGAACAGACAGCCCTCAGAAATAATAGCCGATATCTACAACATAT	29680
Db	895	AspileAspGlnTrpAsnArgThrGluProSerGluileMetProHislleTyrAsnTyr	914
Qy	29679	CTGATCTTTGACAAACCTGAGAAACCAACCAATGGGAAAGGATTCCTTATTAATAAA	29620
Db	915	LeuillePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	934
Qy	29619	TGGTGCCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAACTGGATCCCTTCTTACA	29560
Db	935	TrpCysTrpGluAsnTrpLeuAlaileCysArgLysLeuLysLeuAspLeuPheLeuThr	954
Qy	29559	CCTTATACAAAATAATTCAGATGATTAAAGACTTAAATGTAGACTTAAACCATTA	29500
Db	955	ProTyrThrLyslleAsnSerArgTrpLleLysAspLeuAsnValLysProLysThrille	974
Qy	29499	AAATCCCTAGAGAAACCCAGGCAATACCATTCAGACATAGCATGGGCAAGGACTTC	29440
Db	975	LysThrLeuGluGluAsnLeuGlylleThrIleGlnAspIleglyValGlyLysAspPhe	994
Qy	29439	ATGTCTAAACACCAACCAAGCAATGGCAACAAACCAAAATTCACAAATGGGATCTAAT	29380
Db	995	MetSerLysThrProLysAlaMetAlaThrLysAspLyslleAspLysTrpAspLeuille	1014
Qy	29379	AAACTAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGAGTGAACGCAACTTACA	29320
Db	1015	LysLeuLysSerPheCysThrAlaLysGluThrIleargValasnarGlnProThr	1034
Qy	29319	GAATGGAGAAAATTTTGGCACTACTCATTTGACAAAGGGCTAATATCCAGAACTAC	29260
Db	1035	ThrTrpGluLysllePheAlaThrTyrSerAspLysGlyLeuilleSerArgilleTyr	1054
Qy	29259	AATGAACCTCAACAAATTTACAGAAACCAACCAAGCAACCCCAATCAAAAGTGGGTGAG	29200
Db	1055	AsnGluLeuLysGlnlleTyrLysLysThrAsnAsnProlleLysLysTrpAlaLys	1074
Qy	29199	GATATGAACAGACACTTCTCAAAAGAGGCTTTATGCAGCCCAAAACACATGAAAAA	29140
Db	1075	AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLys	1094
Qy	29139	TACTCATCATCTGGCCATCAGAGAAATGCAATCAAAACCAACATGAGATACCATCTC	29080
Db	1095	CysSerSerLeuAlaileArgGluMetGlnlleLysThrThrMetArgTyrHisLeu	1114
Qy	29079	ACACCACTAGATGGCAATCATTAAAGTCAGGAACCAACAGGCTGGAGAGATGT	29020
Db	1115	ThrProValargMetAlailelleLysLysSerGlyAsnArgCysTrpArgGlyCys	1134
Qy	29019	GGAAATAGGAACACTTTTACACTGTGGTGGGACTGTGAACCTAGTTCACCATTTGG	28960
Db	1135	GlyGluilleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp	1154
Qy	28959	AAGTCAGTGGGTATTCCTCAGGGATCTAGAACTAGAAATACCATTTGACCCAGCCATC	28900
Db	1155	LysSerValTrpArgPheLeuArgAspLeuGluLeuLeuilleProPheAspProAlaile	1174
Qy	28899	CCATTACTGGGTATATACCAAGGATTAATATCATGCTGTATATAAGACACACAGCACA	28840
Db	1175	ProLeuLeuGlylleTyrProGluAspTyrLysSerCysCysTyrLysAspThrCysThr	1194
Qy	28839	TGTATGTTTATAGCAGCACTATTCACAATAGCAAGACTTGAACCAACCTAAATGTCCA	28780
Db	1195	ArgMetPheilleAlaLeuPheThrilleAlaLysThrTrpAsnGlnProLysCysPro	1214
Qy	28779	ACAACGATAGCTGGATTAAGAAAATGTGGCACATATACCATGGAATPACTATGAGCC	28720
Db	1215	ThrMetlleAspTrpIleLysLysMetTrpHislleTyrThrMetGluTyrTyrAlaAla	1234
Qy	28719	ATAAAAATGATGAGTTCGTCTCTTTGTAGGACATGGATGAAGCTGGAACCATCAT	28660
Db	1235	lleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrille	1254
Qy	28659	CTCAGCAACATATCACAGGACAAAAACCAACCAACCGCATGTTCTCACTCATAGGTGG	28600
Db	1255	LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgllePheSerLeuilleGly	1274
Qy	28599	RAT 28597	
Db	1275	Asn 1275	
RESULT 12			
B28096			
line-1 protein ORF2 - human			
C:Species: Homo sapiens (man)			
C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 30-Sep-1993			
C:Accession: B28096			
R:Skowronski, J.; Fanning, T.G.; Singer, M.F.			
Mol. Cell. Biol. 8, 1385-1397, 1988			
A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.			
A:Reference number: A28096; MUID:88246405; PMID:2454389			
A:Accession: B28096			
A:Status: preliminary; not compared with conceptual translation			
A:Molecule type: mRNA			
A:Residues: 1-1275 <SKO>			
C:Superfamily: pol polyprotein			
Alignment Scores:			
Pred. No.: 1.27e-165 Length: 1275			
Score: 2196.00 Matches: 405			
Percent Similarity: 94.10% Conservative: 10			
Best Local Similarity: 91.84% Mismatches: 26			
Query Match: 4.19% Indels: 0			
DB: 2 Gaps: 0			
US-10-083-853B-2 (1-29921) x B28096 (1-1275)			
Qy	29919	ACTTCTTCACAGAAATTTGGAAAACTACTTTAAAGTTCATGTGGAACCAAAAAAGAGCC	29860
Db	835	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla	854
Qy	29859	CACATCGCCAGTCAATCCTTAAGCCAAAGCAAAAGCTGGAGCATCATCTACCTGAC	29800
Db	855	ArgilleAlaLysSerlleLeuSerGlnLysAsnLysAlaGlylleThrLeuProAsp	874
Qy	29799	TTCAAACTACACTACAGGCTACAGTAACCAAAACCAACATGGTACTGGTACCAAAACAGA	29740
Db	875	PhelysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTyrTrpGlnAsnArg	894
Qy	29739	GATATCAACCAATGGAACAGACAGCCCTCAGAAATAATGCCGATATCTACAACATAT	29680
Db	895	AspileAspGlnTrpAsnArgThrGluProSerGluilleThr***LeuThrTyrAsnTyr	914
Qy	29679	CTGATCTTTGACAAACCTGAGAAACCAACCAATGGGAAAGGATTCCTTATTAATAAA	29620
Db	915	LeuillePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys	934
Qy	29619	TGGTGCCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAACTGGATCCCTTCTTACA	29560
Db	935	TrpCysTrpGluAsnTrpLeuAlaileCysArgLysLeuLysVal	

```
|||||
Db 995 MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrAspLeuIle 1014
QY 29379 AAACATAAGAGCTCTGCACAGCAAAAGAAATACCACATCAGAGTCAAGCAACTTACA 29320
Db 1015 LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgProThr 1034
QY 29319 GAATGGGAGAAAATTTTGCACCTACTCTATCTTGCACAAAGGCTTAATATCCAGAACTTAC 29260
Db 1035 ThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 1054
QY 29259 AATGAATCAACAAATTTTACAGAAAAAACAAGAACCCCATCAAAAGTGGGTGAAG 29200
Db 1055 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTyrAlaLys 1074
QY 29199 GATATGAACACACTCTCTCAAAAGAGGCAATTTATGCAGCCAAACACATCAAAAAA 29140
Db 1075 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 1094
QY 29139 TACTCATCATCAGTGGCCATCAGAGAAATGCAAAATCAAAACCAACATGAGATACCATCTC 29080
Db 1095 CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 1114
QY 29079 ACACAGTTAGATGGCAATCATTAAGATCAGGAACAACAGGTGCTGGAGAGAATGT 29020
Db 1115 ThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTyrArgGlyCys 1134
QY 29019 GGAGAAATAGAACACATTTTACACTGTTGGGAGCTGTGAACCTAGTTCACACCATTTGGG 28960
Db 1135 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTyr 1154
QY 28959 AAGTCAGTGTGGTGAATTCCTCAGGAGATCTAGAACTAGAAATPACCATTTGACCCAGCCATC 28900
Db 1155 LysSerValTyrPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174
QY 28899 CCATTACTGGTATATACCCAAAGGATATATAATCATCTGCTCTATAAGACACAAAGCACA 28840
Db 1175 ProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCysTyrLysAspThrCysThr 1194
QY 28839 TGTATGTTATAGCAGCACTATTCCACATAGCAAGACTTGGAAACCACTTAATATGCCA 28780
Db 1195 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1214
QY 28779 ACAACGATAGACTGGATTAAGAAAATGTGGCACATATACACCATGGAATACTATGCAGCC 28720
Db 1215 ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 1234
QY 28719 ATAAAAATGATGATGCTGCTCTTTGTAGGACATGGATGAAGCTGGAAACCATCATTT 28660
Db 1235 IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254
QY 28659 CTCAGCAACATCATCAGAGGACAAAACCAACCAACCGCATGTTCTCATCATAGGTGGG 28600
Db 1255 LeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGlyGly 1274
QY 28599 AAT 28597
Db 1275 Asn 1275
```

## RESULT 13

S23650  
retrovirus-related hypothetical protein II - human retrotransposon LINE-1  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 01-Nov-1996 #text\_change 08-Jan-1999

C:Accession: S23650

R:Hohjoh, H.; Minakami, R.; Sakaki, Y.

Nucleic Acids Res. 18, 4099-4104, 1990

A:Title: Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which

C:Reference number: S23649; MUID:90332398; PMID:2165587

A:Accession: S23650

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-712 <HOH>

A:Cross-references: EMBL:X52235  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1990  
C:Genetics:  
A:Mobile element: LINE-1  
A:Start codon: GTG  
C:Superfamily: pol polyprotein

## Alignment Scores:

Pred. No.:	1.14e-164	Length:	712
Score:	2183.00	Matches:	403
Percent Similarity:	93.88%	Conservative:	11
Best Local Similarity:	91.38%	Mismatches:	27
Query Match:	4.17%	Indels:	0
DB:	2	Gaps:	0

US-10-083-853B-2 (1-29921) x S23650 (1-712)

```
QY 29919 ACTTCTTCACAGAAATTTGAAAAAACTACTTTAAAGTTTCATGTGGAACCAAAAAAGAGCC 29860
Db 272 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTyrAsnGlnLysArgAla 291
QY 29859 CACATCGCCCAAGTCAATCTTAAGCCAAAAAGAAAGCTGGAGGCATCATCTACCTGAC 29800
Db 292 ArgIleAlaLysAlaIleArgSerGlnLysAsnLysSerGlyGlyIleThrLeuProAsp 311
QY 29799 TTCAAACTACATACAGGCTACAGTAAACCAACAAACATGGTACTGTGTACCAAAACAGA 29740
Db 312 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTyrTyrGlnAsnArg 331
QY 29739 GATATCAACCAATGGACAGAACAGAGCCCTCAGAAATTAATGTTAGACCTAAACCATTA 29680
Db 332 AspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProHisIleTyrAsnTyr 351
QY 29679 CTGATCTTTGACAAACCTGAGAAAACAAGCAATGGGAAAGGATTCCTCTATTAAATAAA 29620
Db 352 LeuIlePheAspLysProGluLysAsnGluGlnTrpGlyLysAspSerLeuPheAsnLys 371
QY 29619 TGGTCTGAGAAAAAATCGCTAGCCCTATGTAGAAGCTGAAACCTGGATCCCTCTTCCATA 29560
Db 372 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 391
QY 29559 CCTATACAAAATTAATTCAGATGGATTAAAGACTTAATGTTAGACCTAAACCATTA 29500
Db 392 ProTyrThrLysIleAsnSerArgTyrIleLysAspLeuIleValArgProLysThrIle 411
QY 29499 AAATCCCTAGAGAAAACCCAGCAATACCATTCAGGACATAGGCATGGCAAGGACTTC 29440
Db 412 LysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPhe 431
QY 29439 ATGTCTAAACACCAAAAGCAATGGCAACAAAGCCAAAATTTGACAAAATGGGATCTAAT 29380
Db 432 MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrAspLeuIle 451
QY 29379 AAATAAGAGCTTCTGCACAGCAAAAAGAACTACCATCAGAGTGAACAGGCAACTTACA 29320
Db 452 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 471
QY 29319 GAATGGGAGAAAATTTTGGCAACCTACTCTATTGACAAAGGCTTAATATCCAGAACTTAC 29260
Db 472 LysTrpGluLysThrPheAlaThrTyrSerSerAspLysGlyLeuIleSerArgIleTyr 491
QY 29259 AATGAATCAACAAATTTTACAGAAAAAACAAGAACCCCATCAAAAGTGGGTGAAG 29200
Db 492 AsnGluLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTyrAlaLys 511
QY 29199 GATATGAACACACTTCTCAAAAAGAGGCAATTTATGCAGCCAAACACATCAAAAAA 29140
Db 512 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaLysLysHisMetLysLys 531
QY 29139 TACTCATCATCAGTGGCCATCAGAGAAATGCAAAATCAAAACCAACATGAGATACCATCTC 29080
Db 532 CysSerProSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 551
```



```

QY 29079 ACACAGCTTGAATGGCAATCATATAAAAGTCAGGAAACACAGGCTGCTGGAGAGAAATGT 29020
Db |||
552 ThrProValArgMetAlaIleLeuLysLysSerGlyAsnAsnArgCysTsrPArgGlyCys 571
QY 29019 GGAGAAATAGGACACTTTTACACTGTTGGTGGGACTGTCAACTAGTTTCAACCAATGTGG 28960
Db |||
572 GlyGluIleGlyThrLeuLeuHisCysTsrTrpAspCysLysLeuValGlnProLeuTrp 591
QY 28959 AAGTCAGTGTGGTGATTCCTCAGGGATCTAGAACTAGAAATACCATTTGACCCAGCCATC 28900
Db |||
592 LysSerValTrpArgPheLeuArgAspLeuGluProGluIleProPheAspProAlaIle 611
QY 28899 CCATTACTGGGTATATACCAAGGATTATAAATCATGCTGCTATAAAGACACAAAGCACA 28840
Db |||
612 ProLeuLeuGlyIleValProLysAspSerLysSerCysCysTyrLysAspThrCysThr 631
QY 28839 TGTATGTTTATACAGCACTATTTCACAATAGCAAGACACTGGAAACCAACCTTAATGTCCA 28780
Db |||
632 ArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 651
QY 28779 ACAACGATAGACTGGATTAGAAAATGTGGCACATATACACCATGGAATACTATGCAGCC 28720
Db |||
652 ThrMetIleAspTrpIleLysLysMetTrpHisIleThrMetGluTyrTrpAlaAla 671
QY 28719 ATAAAAATGATGAGTTGGTCTCTTTGTAGGACATGGATGAAGCTGGAACCATCATTT 28660
Db |||
672 IleLysAsnAspGluPheValSerPheValGlyThrTrpMetLysLeuGluIleIle 691
QY 28659 CTGAGCAAACTATCACAGGACAAAAACCAACACCGCATGTTCTCACTCATAGTGGG 28600
Db |||
692 LeuSerLysLeuSerGlnGluGlnLysThrLysHisCysIlePheSerLeuIleGlyGly 711
QY 28599 AAT 28597
Db |||
712 Asn 712

RESULT 14
JU0033
hypothetical L1 protein (third intron of gene TS) - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 13-Sep-1996
C:Accession: JU0033
R:Horie, N.; Naibantoglu, J.; Kaneda, S.; Ayusawa, D.; Seno, T.; Takeishi, K.
J. Biochem. 106, 1-4, 1989
A:Title: Identification and characterization of an L1 family sequence with a very long
A:Reference number: JU0033; PMID:89380111; PMID:2476429
A:Accession: JU0033
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-562 <HOR>
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 3,49e-164 Length: 562
Score: 2176.50 Matches: 449
Percent Similarity: 85.82% Conservative: 35
Best Local Similarity: 79.61% Mismatches: 74
Query Match: 4.16% Indels: 10
DB: 2 Gaps: 2

US-10-083-853b-2 (1-29921) x JU0033 (1-562)

QY 5429 CAGCATCATCTCGATACCAAAACCTAGCAGTGACACACAAAAAGAGAAATTTTCAGGCC 5488
Db |||
2 GlnHisHisProAspThrLysAlaGlyGlnArgHis-ThrLysLysGluAsnPheArgPr 21
QY 5489 CATATCCCTGATGAACATTGATGTGAAATCTCAATAAATCTCAATAAATCTGGCAAAACCAATCCA 5548
Db |||
21 IleSerLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleG1 41
QY 5549 GCAGCACATCAAAAGCTTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATGCAAGG 5608
Db |||

```

```

41 nGlnHisIleLysLysLeuIleHisAspHisValGlyPheIleProArgMetGlnG1 61
5609 CTGGTTCAAAATATGCAAAATCAATAATGTAGGCCATCATATAACAGAACCAATGACAA 5668
Db |||
61 YTrpPheAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgSerLysAspLy 81
5669 AAACCAACATGATTATTCATATAGATGCAGAAAAAGCCTTTGTGTCACAAATTCACACAGCCCTT 5728
Db |||
81 sAsnHisMetIleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnHisI1 101
5729 CATGCTAAATAATTCCTCAGTAACTAGTATCGATGGAATGTATCTCAAAATATAAGAGC 5788
Db |||
101 eMetLeuLysThrLeuAsnLysLeuGlyIleAspGlyThrTyrPheLysIleIleArgAl 121
5789 TATTTATAC-AAACCCACACAGCCAAATATCATCTACTGAATGGCAAAAACTTGAAGCATTCCTCC 5847
Db |||
121 alleTyrAspLysAlaThrAlaAsnIleIleValAsnGlyGlnLysLeuGluAlaPhePr 141
5848 TTTGAGAAGTGCACAAAGACAGGATGCCCTCTCTCACCACTCCTATTTCAAGATACTATT 5907
Db |||
141 oLeuLysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValle 161
5908 GGAACTTCGGCCAGGCAATCAGGCAATAGAAAGAAATAAAGGGTATTCAATAGTAAAG 5967
Db |||
161 uGluValLeuAlaArgAlaIleArgGlnLysGluIleLysGlyIleGlnLeuGlyLy 181
5968 AGAGGAAGTCATATTGCTCTGTTGCAGATGACATGTTGTATATTATAGAAAACCCCAT 6027
Db |||
181 sGluGluValLysLeuSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProI1 201
6028 CGTCTCAGGCCCAAAATCTCTTAAGCTGATAGCAAACTTCAGCAAAAGTCTCAGGACACAA 6087
Db |||
201 eValSerAlaGlnAsnLeuLysLeuIleSerSerPheSerLysValSerGlyTyrLy 221
6088 AATCAATGTCRAAAATCACAACCATTCCTATACGCCCAATATAGACAAACAGAGACCA 6147
Db |||
221 sIleAsnLeuGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerG1 241
6148 AATCATGAGTGAACCTCTCATTCACAATTCGTACAAAGAGAATAAATACTAGGAATACA 6207
Db |||
241 nIleMetSerGluLeuProLeuThrIleAlaSerLysArgIleLysTyrLeuGlyIleH1 261
6208 ACTTACAGGGACACCTAGGAACCTCTTCAAGGAGAACTCAAAACCACTGATCAAGGAAT 6267
Db |||
261 sLeuLysArgAspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuAsnGluI1 281
6268 AAGAGAGGACACAAACAAATGGAACCATTCCTCCTCACAGATAGTAAAGATCAT--- 6324
Db |||
281 eLysGluAspThrLysLysTrpLysThrIleProCysSerTrpValGlyArgIleAsnI1 301
6325 ---GAAAATGCCATCTGCCAAAGTAAATATAGATTCAGTGTCTACCCCTCAAGCT 6380
Db |||
301 eValLysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLe 321
6381 ACATTGACTTTCTTCACAGAATTTGAAAAACAACTTTAAATTTTCAATTTGGAACCAAA 6440
Db |||
321 uProMetThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-L 341
6441 AAGAGCCACACAGACCAATCTTAAGCAAAAGAAAGCAAGCTGGAGGTATCATGC 6500
Db |||
341 ysArgAlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrL 361
6501 TACCTGACTTAAACTATATAGGCTACAGTACCAAACTACCAAACTGCTACTGTGTACC 6560
Db |||
361 euProAspPheLysLeuTyrTyrLysAlaThrValThrLysThrValTyrTyrTrpTyG 381
6561 AAACAGATATATAGACCAATGGACAGAACACAGACCTCAGAAATTCAGACT-GCAATCT 6619
Db |||
381 InAsnArgHisIleAspGlnTrpAsnArgThrGluProSerGluIleThrProHisIleT 401
6620 ACATCCATCTGATCTTTGACAAACCTTGACAAACCAAGCAAGCAAGTGGAAAGGATTCCTCAT 6679
Db |||
401 yrasnTyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuc 421

```





Db 1254 LeuSerLysLeuSerGln 1259

Search completed: July 4, 2003, 22:37:08  
Job time : 785 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 4, 2003, 20:41:28 ; Search time 515.5 Seconds

(without alignments)

13348.993 Million cell updates/sec

Title: US-10-083-853b-2  
Refseq scores: 52328  
Sequence: 1 gatatgtgaagaagctca.....caattctgtaagaagtaa 29921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 881726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPRO.spool/US10083853/runat\_03072003\_093613\_8176/app\_query.fasta\_1.30087  
-DB=publishedApplications\_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10083853 @CGN\_1\_1\_192\_runat\_03072003\_093613\_8176  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database 8 Published Applications Archive

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5169	9.9	1275	US-10-025-201-3
2	4081.5	7.8	1010	US-10-050-882-80
3	3770.5	6.9	956	US-10-000-256A-153
4	3628.5	6.9	917	US-10-001-835-189

5	3535.5	6.8	940	9	US-10-082-830-235	Sequence 235, App
6	2215	4.2	1275	9	US-10-025-201-3	Sequence 3, Appli
7	2199	4.2	1010	9	US-10-050-882-80	Sequence 80, Appl
8	2187	4.2	956	9	US-10-000-256A-153	Sequence 153, App
9	1467	2.8	917	9	US-10-001-835-189	Sequence 189, App
10	1351	2.6	940	9	US-10-082-830-235	Sequence 235, App
11	1090.5	2.1	338	9	US-10-025-201-2	Sequence 2, Appli
12	806	2.1	317	9	US-10-083-853-2	Sequence 2, Appli
13	767.5	1.5	341	9	US-10-001-876-169	Sequence 169, App
14	708	1.4	492	9	US-10-007-280A-212	Sequence 212, App
15	676	1.3	190	9	US-10-074-475-250	Sequence 250, App
16	630	1.2	492	9	US-10-007-280A-212	Sequence 212, App
17	590.5	1.1	315	12	US-10-001-843-125	Sequence 125, App
18	580	1.1	192	9	US-10-001-857-119	Sequence 119, App
19	573.5	1.1	317	9	US-10-083-853-2	Sequence 2, Appli
20	548.5	1.0	197	9	US-10-157-031-367	Sequence 367, App
21	533.5	1.0	192	9	US-10-001-857-119	Sequence 119, App
22	533	1.0	315	12	US-10-001-843-125	Sequence 125, App
23	481.5	0.9	212	10	US-09-864-761-46654	Sequence 46654, A
24	460.5	0.9	230	9	US-10-000-256A-143	Sequence 143, App
25	434	0.8	113	9	US-10-012-542-399	Sequence 399, App
26	431.5	0.8	316	12	US-10-001-870-132	Sequence 132, App
27	429.5	0.8	121	9	US-10-106-698-4424	Sequence 4424, Ap
28	429	0.8	167	10	US-09-864-761-34496	Sequence 34496, A
29	418	0.8	101	9	US-10-012-542-400	Sequence 400, App
30	414.5	0.8	121	9	US-10-106-698-4424	Sequence 4424, Ap
31	409	0.8	322	9	US-10-086-623-6	Sequence 6, Appli
32	409	0.8	322	9	US-10-260-539-6	Sequence 6, Appli
33	409	0.8	364	9	US-10-028-072-186	Sequence 186, App
34	409	0.8	364	9	US-10-121-049-186	Sequence 186, App
35	409	0.8	364	9	US-10-123-904-186	Sequence 186, App
36	409	0.8	364	9	US-10-140-470-186	Sequence 186, App
37	409	0.8	364	9	US-10-175-746-186	Sequence 186, App
38	409	0.8	364	9	US-10-176-918-186	Sequence 186, App
39	409	0.8	364	9	US-10-176-921-186	Sequence 186, App
40	409	0.8	364	9	US-10-137-865-186	Sequence 186, App
41	409	0.8	364	9	US-10-140-474-186	Sequence 186, App
42	409	0.8	364	9	US-10-142-431-186	Sequence 186, App
43	409	0.8	364	9	US-10-143-114-186	Sequence 186, App
44	409	0.8	364	9	US-10-140-002-186	Sequence 186, App
45	409	0.8	364	9	US-10-142-419-186	Sequence 186, App

ALIGNMENTS

RESULT 1  
US-10-025-201-3  
; Sequence 3, Application US/10025201  
; Publication No. US20030003468A1  
; GENERAL INFORMATION:  
; APPLICANT: Crow, Mary K.  
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY  
; FILE REFERENCE: 5983/2H567  
; CURRENT APPLICATION NUMBER: US/10/025,201  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,673  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 3  
; LENGTH: 1275  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U09116  
; DATABASE ENTRY DATE: 1995-02-02  
; RELEVANT RESIDUES: (1)..(1275)  
US-10-025-201-3

Alignment Scores:  
Pred. No.: 0  
Score: 5169.00  
Percent Similarity: 78.90%  
Length: 1275  
Matches: 1054  
Conservative: 79

Best Local Similarity: 73.40% Mismatches: 137  
Query Match: 9.88% Indels: 173  
DB: 9 Gaps: 7

US-10-083-853B-2 (1-29921) x US-10-025-201-3 (1-1275)

QY 3428 ATGGCAGGATCAATATTCACACATATAATATACTTAATTAATGTAATGGCTAAATTC 3487  
DB 1 MetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnIleAsnGlyLeuAsnSer 20  
QY 3488 CCAATTAAGACACAGACTGGCAATTTGGATAAGAGTCAAGACCCATCAGTGTGCTGT 3547  
DB 21 AlaIleLysArgHisArgLeuAlaSerTrpIleLysSerGlnAspProSerValCysCys 40  
QY 3548 ATTGAGAGGCCCTCTCACATGAAAGACACACATAGGCTCAAAATAAGGGATGGAGG 3607  
DB 41 IleGlnGluThrHisLeuMetCysArgAspThrHisArgLeuLysIleLysGlyTrpArg 60  
QY 3608 AAGATTACCAAGTAATGGAAACACAAAAAAGCAGGGGTGGCAATCCTAGTCTCT 3667  
DB 61 LysIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyValAlaIleLeuValSer 79  
QY 3668 GATAAAGACAGCTTTAAACCAACAAGATCAAAAGACACAAAGAGGCCATTACATAATG 3727  
DB 80 AspThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrIleMet 99  
QY 3728 GTAAGGCATCAATGGACAAAGAGAGTAACTACTCTTAATATACATGACCCCAATACA 3787  
DB 100 ValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThr 119  
QY 3788 GGACACCCAGATTCATAAAGCAAGTCTTAGACACCTACAAAGAGACTTTGACTCCAC 3847  
DB 120 GlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAspSerHis 139  
QY 3848 ACAATAATAGTGGGAGTCTAAATAATAATAGACACTTTAAACCCCACTGCCAATATTA 3907  
DB 140 ThrLeuIleMetGly-----AspPheAsnThrProLeuSerIleLeu 153  
QY 3908 GGCAGATCAATGACAGACAAATTAACAAGGATATCCAGGAGTTGAACCTGAGCTCTGCAC 3967  
DB 154 AspArgSerThrArgGlnLysValAsnLysAspThrGlnLeuLeuAsnSerAlaLeuHis 173  
QY 3968 CAAGCGACCTAATAGATATCTACAGACTCCCAACCCCAATCAACAGATATACACTC 4027  
DB 174 GlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSerThrGluTyrThrPhe 193  
QY 4028 TTCTCAGCATCACATTAACCATTTAAATTCACCATGTAATTTAAAGTAAACCACTC 4087  
DB 194 PheSerAlaProHisHisThrTyrSerLysIleAspHisIleValGlySerLysAlaLeu 213  
QY 4088 CTCAGCAATCAAAAGACAGAAATCTCTAACAACAGTCTCTCAGACTACAGTCAATC 4147  
DB 214 LeuSerLysCysLysArgThrGluIleThrAsnTyrLeuSerAspHisSerAlaIle 233  
QY 4148 TATTAGACCTCAAAATTAAGAACTCACTCAAAATCACAACTACATGGAACATCAAC 4207  
DB 234 LysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTrpLysLeuAsn 253  
QY 4208 AACCTGCTCCTGAATGACTACTGGTAAATAAACAATAAGGCAAAATAAAGATGTC 4267  
DB 254 AsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAlaGluIleLysMetPhe 273  
QY 4268 TTTGAAACCAATGAGACAAAGACACAAATGTACAGAACTCTCTGGGCATATTTAAAGCA 4327  
DB 274 PheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPheLysAla 293  
QY 4328 GTGTGTAGAGGGAATTTATAGACTAGTCCCTACAGAGAAAGAGGAAATCTAAA 4387  
DB 294 ValCysArgGlyLysPheIleAlaLeuAsnAlaTyrLysArgLysGlnGluArgSerLys 313  
QY 4388 ATACACACCTTAACATCACATTAAGAACTACAGAGAAAGAGCAACAAATTCAAA 4447  
DB 314 IleAspThrLeuThrSerGlnLeuLysGluLeuLysGlnGluGlnThrHisSerLys 333

QY 4448 GCTAGCAGACAGACAAAGAAATAACTAAGATCAGACAGAACTGAAGAGATAGACACAA 4507  
DB 334 AlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGluThrGln 353  
QY 4508 AAGCCCTTCAATAAATAATCAATGAATCCAGAGAGCTGTTTTTTTCAAAAGATCAGCAAAAT 4567  
DB 354 LysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePheGluArgIleAsnLysI 373  
QY 4568 ----AGACCACTAGACAGACTAATAAAGAAAGAGAGAGAAATCAAAAGAGATGCAAT 4623  
DB 373 eAspArgProLeuAlaArgLeuIleLysLysArgLysLysAsnGlnIleAspThrI 393  
QY 4624 AAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATCAAAACTATTATCAGAGA 4683  
DB 393 eLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIleArgG 413  
QY 4684 ATATTATAACACCTCTATGCAATAATAACTAGAAAACTTAGAGAAATGATTAATTCCT 4743  
DB 413 uTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluMetAspThrPheLe 433  
QY 4744 GGACACATATGTAGCCTGTATGGACCTTGGGGGACAGACAAAGGGGTGAATGCAGAA 4803  
DB 433 uAsp----- 434  
QY 4804 ATAAAGACAAAGACAAAGAGTATGTTTGGAAAGTAGGGTTCAGGGGGCAACTTGCCTCT 4863  
DB 434 ----- 434  
QY 4864 AATGGACAAAGGGCCCTGAGCTTTACACCACTCTGTATTATTAGCAAAAGAGATAGC 4923  
DB 434 ----- 434  
QY 4924 GAGAGGCTGAGTTGGAAGAGAGAGTCACTGTAGGTCCAGAGTAGGCTGCAAGACTGC 4983  
DB 434 ----- 434  
QY 4984 ATTCTCAACAATAGGCTCTAGATGTCACAGTAGATAACCTCAAGGAGCCAGTGCAGG 5043  
DB 434 ----- 434  
QY 5044 GAGTGATGGCCCTCAGCAAACTTCTAGGGCAGCAGACAGAAGTAAGTTTGCCTCATTCT 5103  
DB 434 ----- 434  
QY 5104 GTATTACGATAAAGACTTTGCTGTTTGTATCAAGTAGCTCCAGTGAATGCTGAGTTGG 5163  
DB 434 ----- 434  
QY 5164 TCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTAAACCA 5223  
DB 435 -----ThrTyrThrLeuProArgLeuLysG 443  
QY 5224 GGAGAAGTCAAAATCCCTGAATATACAGTAACAGAGTTCTAAATTTCAACGCAATTAATGA 5283  
DB 443 nGluGluValGluSerLeuAsnGlyProIleThrGlySerGluIleValAlaIleIleAs 463  
QY 5284 TAGCTACCAACCAAAAAAGTCCAGGACAGAGGATTCACAGCCAAATTTCTACACAG 5343  
DB 463 nSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPheTyrGlnAr 483  
QY 5344 GTACAAAGAGAGCTGGTACTATTCCTTCTGAAACTATTCCAAAAATAGAA---AATGG 5400  
DB 483 gTyrLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-G 503  
QY 5401 GAATCTCCCTTAACCTATTTTACAGGCCAGCATCATCTCTGATACCAAAACCTAGCAGTG 5460  
DB 503 lTyrLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysProGlyArg 523  
QY 5461 ACACAACAAAGAGAGAAATTTTCAAGGCCCATATCTCCCTGTAGCAATGTGTGAAATTC 5520  
DB 523 spThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAlaLysIleL 543

```
QY 5521 TCAATATAAATACTGGCAAAACCAATCCAGCAGCACATCAAAAGACTTATCTACCATGATC 5580
Db 543 euAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHisIleHisaspG 563
QY 5581 AAGTTGGCGTCATCCCTGGGATGCAAGCTGGTTCAAAATATGCAAAATCAATAATGTAG 5640
Db 563 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleAArgLysSerIleAsnValI 583
QY 5641 GCCATCATATAAAGACCAATGACAAACCAACCATGATTTATCTCAATAGATGCAAAA 5700
Db 583 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 603
QY 5701 AGGCTTTGTCAAAATTAACAGACCCCTTCATGCTGATAAAATCTCAGTAAACTAGGTATCG 5760
Db 603 ysaAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 623
QY 5761 ATGGAATGTATCTCAAAATATAAGACTATTTATAC- AAACCCACACGCCCAATATCATAC 5819
Db 623 spGlyThrTyPheLysIleIleAArgAlaIleTyAspLysProThrAlaAsnIleIleL 643
QY 5820 TGAATGGCAAAACTGGAAGCATTCCTTTGAGAAGCTGGCACAGACAGGATGCCCTC 5879
Db 643 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 663
QY 5880 TCTCACACACTCTATTAAGATATAATTGGAAAGTTCTGGCCAGGCAATCAGGCAATAGA 5939
Db 663 euSerProLeuPheAsnIleValLeuGluValLeuAlaAArgAlaIleAArgGlnGluL 683
QY 5940 AAGAAATAAGGGTATCAATAGAAAGAGAGAGAGTCAATATCTCTCTGTTTCCAGATG 5999
Db 683 ysgLuiIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 703
QY 6000 ACATGTTTGTATATTAGAAAACCCCTGCTCAGGCCCAAAACTCCTTAAGCTGATAA 6059
Db 703 spMetIleValTyLeuGluAsnProIleValSerAlaGlnAsnLeuLysLeuIleS 723
QY 6060 GCAACTTCAGAAAGTCTCAGGACACAAATCAATGTCGAAAATCAAGCATCTTAT 6119
Db 723 erAsnPheSerLysValSerGlyTyLysIleAsnValGlnLysSerGlnAlaPheLeuT 743
QY 6120 ACGCAATAATAGACAAACAGAGACCAAAATCATGAGTGAACCTCTCATCTCACAATTGCTA 6179
Db 743 yrThrAsnAsnArgGlnThrGluSerGlnIleMetGlyGluLeuProPheThrIleAlaS 763
QY 6180 CAAGAGAATAAATACTAGGAATACAACCTTACAAGGGACACGCTAGGAACTCTTCAAGG 6239
Db 763 erLysArgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 783
QY 6240 AGAACTCAAAACCACTGATCAAGGAATAAGAGAGGACACAAACAAATGGAAAACATTC 6299
Db 783 luAsnTyTyLysProLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP 803
QY 6300 CATGCTCACAGATAGTAAGATCAT-----GAAATGGCATACTGCCCAAGCAATATT 6352
Db 803 roCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIleT 823
QY 6353 ATAGATTCAAGTGTACCCCTCAAGCTACCATTTGACTTCTTCCACAGAATTTGAAAAA 6412
Db 823 yrArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLysT 843
QY 6413 CAACCTTAATTTTCATATGGACCAAAAGAGCCCAAGAGCCACAGACCAATCTTAAGC 6472
Db 843 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaAArgIleAlaLysSerIleLeuSer 862
QY 6473 AAAAGAACAAAGCTGAGGTATCATCTACCTACCTTAAACTATATACTATAAGGCTACA 6532
Db 863 GlnLysAsnLysAlaGlyGlyIleThrLeuProTyPheLysLeuTyTyLysAlaThr 882
QY 6533 GTAACCAAAAGCTGCTACTGTTACCAAAACAGATATATAGACCAATGGAACAGAAC 6592
Db 883 ValThrLysThrAlaTrpTyTyTrpTyGlnAsnArgAspIleAspGlnTrpAsnArgThr 902
QY 6593 GAGACCTCAGAAATTAC- ACTGCAATCTACATCATCTGATCTTTTGACAAACCTGACAAA 6651
```

```
Db 903 GluProSerGluIleMetProHisIleTyAsnTyTrLeuIlePheAspLysProGluLys 922
QY 6652 AACAAAGCAATGAAAAAGGATTCCCTATTTTAAATAAATGGTCTTGGAAGAACTGGCTAGCC 6711
Db 923 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla 942
QY 6712 ATATGCAAAAGCTGAAACTGGATCCCTTCCATCCCTTATACACCTTATACAAAAGTTAACTCAAGA 6771
Db 943 IleCysArgLysLeuLysLeuAspLeuPheLeuThrProTyTrThrLysIleAsnSerArg 962
QY 6772 TGAATTAAGACTTAAATATAAGACATATAAACCAATATAAACCCCA- GAAGAAAACTTAGCC 6830
Db 963 TrpIleLysAspLeuAsnValLysProLysThrIleLysThrLeuGluGluAsnLeuGly 982
QY 6831 AATACCATTACAGGATATGGACATGGCAAGACCTTCATGACTATAACACCAACCAAGCAATG 6890
Db 983 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 1002
QY 6891 GCAACAAAAGCAAAATAGACAGCTGGGATCTGATTAACCTATAGAGCTTCTGCACAGCA 6950
Db 1003 AlaThrLysAspLysIleAspLysTrpAspLeuIleLysLeuLysSerPheCysThrAla 1022
QY 6951 AAAAAAAGCTCATCATCAGTGAACCAACCAACCTACAGATGGGAGAGAAATTTTGGCAAT 7010
Db 1023 -LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh 1042
QY 7011 CTATCGATCTGCACAAAGCGCTAATATCCAGAGATCTACGAGAAGCTTAACCAAAATTTACAA 7070
Db 1042 rTySerSerAspLysGlyLeuIleSerArgIleTyAsnGlnLeuLysGlnIleTyLys 1062
QY 7071 GAAAAA----AACAAACCCCTCAAAATATGGGAAAAGATATGAGACAGACTTCTCAAA 7126
Db 1062 LysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerLys 1082
QY 7127 AGAAGACATTTATCGACCCCAACAAACATATGAAAAAACCTCATCATCATTTGGTGGTTAG 7186
Db 1082 sgluAspIleTyAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 1102
QY 7187 AGAATCAAAACAAACCAACAGTGCATACCATCTCATGCTAGTGTAGATGGTGATCAC 7246
Db 1102 gLueMetGlnIleLysThrThrMetArgTyHisLeuThrProValArgMetAlaIleI 1122
QY 7247 TAAAGATCAGGAAACAAACAAATCTGGAGAGGATGGGAGAAATAGGACACTTTTCCA 7306
Db 1122 eLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuH 1142
QY 7307 CTGTTGGTGGGAATGTAATTAGTTCAACCATTTCTGGAAGACAGCTGGGAGATTCCCTTA 7366
Db 1142 sCysTrpTrpAspCysLysLeuValGlnProLeuTrpLysSerValTrpArgPheLeuAr 1162
QY 7367 GGATCTAGAACCAACAAATATCATTTGACCCAGCAATCCCATTTACTGAGTATATATACCAAA 7426
Db 1162 gaspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGlyIleTyProGl 1182
QY 7427 GGAATATAAATCTTATTATTAAGACACATGCACACATATGTTTATTCAGACACTCAT 7486
Db 1182 uAspTyLysSerCysCysTyLysAspThrCysThrArgMetPheIleAlaLeuPh 1202
QY 7487 CACAAATAGCAAAAGCTTGGAAACCAACCAATGTCATCTAGTATAGTACGTGGATTAAGAA 7546
Db 1202 eThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetIleAspTrpIleLysLy 1222
QY 7547 AACATGCAATATACACCATGAATACTACTGAGCCATATAAAG-CATGAGTTTCATGTC 7605
Db 1222 smetTrpHisIleTyThrMetGluTyTyAlaAlaIleLysAsnAspGluPheIleSe 1242
QY 7606 CTTTGCAGAGATATGATGAGCTGGAACCAACCATCATTTCTCAGCAAACTAACACAGAAC 7665
Db 1242 rPheValGlyThrTrpMetLysLeuGluThrIleIleLeuSerLysLeuSerGlnGluGl 1262
QY 7666 GAAACCAACACACCATGTTCTCATTGTAAGTGGAGT 7705
```

Db 1262 nLysThrLysHisArgIlePheSerLeuIleGlyGlyAsn 1275

RESULT 2

US-10-050-882-80  
; Sequence 80, Application US/10050882  
; Publication No. US2003010400A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 27 Human secreted proteins  
; FILE REFERENCE: P2038P1  
; CURRENT APPLICATION NUMBER: US/10/050,882  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: 09/661,453  
; PRIOR FILING DATE: 2000-09-13  
; PRIOR APPLICATION NUMBER: PCT/US00/06783  
; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: 60/125,055  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 156  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 80  
; LENGTH: 1010

TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:

NAME/KEY: SITE  
LOCATION: (25)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (104)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (194)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (362)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (525)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (643)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (649)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (656)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (660)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (731)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (770)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (777)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (790)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (800)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (825)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (987)  
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE  
; LOCATION: (996)  
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (1003)  
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
US-10-050-882-80

Alignment Scores:

Pred. No.: 0 Length: 1010  
Score: 4081.50 Matches: 841  
Percent Similarity: 76.80% Conservatives: 53  
Best Local Similarity: 72.25% Mismatches: 111  
Query Match: 7.80% Indels: 166  
DB: 9 Gaps: 5

US-10-083-853B-2 (1-29921) x US-10-050-882-80 (1-1010)

QY 4244 ATGAAGGCAAAATAAAGATGTTCTTTGAAACCAATGAGACAAAGACACAATGTACCAG 4303  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MetLysAlaGluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyrGln 20  
QY 4304 AATCTCTGGGCATATTAAACGAGTGTCTAGAGGAAATTTATAGCACTAGTGCCTAC 4363  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 21 AsnLeuTrpasp\*\*\*PheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHis 40  
QY 4364 AAGAGAAACGACGAAATATCTAAATAGACACCTTAAACATCAACAATTTAAAGAACTAGAG 4423  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 41 LysArgLysGlnGluArgSerLysIleaspThrLeuThrSerGlnLeuLysGluLeuGlu 60  
QY 4424 AAGAAGAGCAACAAATTCAAAGCTAGCAGACAGCAAGAAATACTAAGATCAGAGCA 4483  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 LysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAla 80  
QY 4484 GAACCTGAAGGAGATAGACACACAAAAGCCCTTCAATAATAATCAATGAATCAGGAGCTG 4543  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 81 GluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTr 100  
QY 4544 GTTTTGTGAAAAGATCAGCAAAAT---AGACCCTAGACAGACTAATAAACAAGAAAG 4599  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 100 pPheGlu\*\*\*IleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLys 120  
QY 4600 AGAGAAGATCAAGAGATGCAATAAATAAATGATAAAGGGGATATCACCCGATCCAC 4659  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 120 gGluLysAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAspProth 140  
QY 4660 AGAATACAACTATTATCAGAGATATTATAAACACCTCTATGCAATATAACTAGAAA 4719  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 140 rGluIleGlnThrThrIleArgGluTyrTrLysHisLeuTyrAlaAsnLysLeuGluAs 160  
QY 4720 TCTAGAAGAAATGGATAAATTCCTGGACACATATGTAGCCTGTATGGACCTTGGGGGACA 4779  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 160 nLeuGluGluMetAspLysPheLeuasp----- 169  
QY 4780 GAACAAAGGGGGTGAATGCAGAAATAAAGACAAAGACAAAGAGATGTGTTTGAAGTA 4839  
Db 169 ----- 169  
QY 4840 GGGGTTCAGGGGCAACTTGCCTCTAATGGACAAAGGCCCTGAGCTTTACACCCTCTG 4899  
Db 169 ----- 169  
QY 4900 TATTTATTAGGCAAAAGAGATAGCGAGGGGTGAGTTGGAGAGAGGTTCAGCTCTTAGG 4959  
Db 169 ----- 169  
QY 4960 TCCAGAGTAGGCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCCAGTAGA 5019  
Db 169 ----- 169  
QY 5020 TAACTCAAGGAGCAGTGCCAGGGAGTGATGGCCCTCAGCAAACTTCTAGGCGAGCA 5079  
Db 169 ----- 169

QY 5080 CAGAAGTAAGTTTGGCCACATCTCTGTATTACGATAAACAGTTTGGCTGTTTGCATCAAGTA 5139  
Db 169 ----- 169  
QY 5140 GCCTCCAGTAGTGGTGGTGCATGATGCCCTTTGGCTTTTGGCTCCCAACAC 5199  
Db 170 -----Th 170  
QY 5200 ATACACCTCTCAGACTAAACAGGAAGTCAATCCCTCAATATACAGTAACAAG 5259  
Db 170 rTyrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGI 190  
QY 5260 TTCTAAATTAAGACGATTAATGATAGCTTACCAACCAAAAGTCCAGGACGACGG 5319  
Db 190 ySerGluile\*\*AlaIleAsnSerLeuProThrLysLysSerProGlyProAspGI 210  
QY 5320 ATTACAGCCAAATTTACAGAGGTACAAAGAGAGCTGGTACTATTCTCTGAACT 5379  
Db 210 yPheThrAlaGluPheTyrGlnArgTyrLysGluLeuValProPheLeuLysLe 230  
QY 5380 ATTCCAAAATAGAA---AATGGGAATCCTCCTAACTCATTTACGAGGCCAGCATCA 5436  
Db 230 uPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIle 250  
QY 5437 TCCTGATACCAAAACCTAGCAGTGACACAAACAAAGAGGAAATTTACGAGGCCATATCCC 5496  
Db 250 leLeuileProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerL 270  
QY 5497 TGATGAACATTGATGTGMAAATCCTCAATAAATCTGCGAAACCAATCCAGCAGACA 5556  
Db 270 euMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisI 290  
QY 5557 TCAAAAGCTTATCTACCATGATCAAGTTGCGTGCATCCCTGGATGCGAAGCTGGTCA 5616  
Db 290 leLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTyrPheA 310  
QY 5617 AAATATGCAATCAATAAATGTAGGCCATCACATAACAGAACCAATGACAAAAACACA 5676  
Db 310 snIleArgLysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnHis 330  
QY 5677 TGATTATCTCAATAGATGAGAAAGCCCTTTGTCAAAATTCACAGCCCTTCATGCTAA 5736  
Db 330 etIleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuL 350  
QY 5737 AAATCTCACTAACTAGGTATCGATGGAATGTATCTCAAAATAAAGAGCTATTATA 5796  
Db 350 yThrLeuAsnLysLeuGlyIleAspGlyThrTyr\*\*LysIleIleArgAlaIleTyrA 370  
QY 5797 C-AAACCCACAGCCAATATCATCTGAATGGCAAAACCTGGAAGCATTCCTTTGAGAA 5855  
Db 370 sPlysProThrAlaAsnIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysT 390  
QY 5856 CTGGCACAACAGAGATGCCCTCTCACCACCTCTATTTCAGATACATATTGGAAGTTC 5915  
Db 390 hrGlyThrArgGlnGlyCysProLeuSerProLeuPheAsnIleValLeuGluVal 410  
QY 5916 TGCCCGGGCAATCAGGCAATAGAAAGAAATAAGGTATTCAATAGAAAGAGAGGAAG 5975  
Db 410 euAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluLuv 430  
QY 5976 TCATATGCTCTGTTTGCAGATGACATGTTTGTATATTATTAGAAAACCCCATGCTCTCAG 6035  
Db 430 allLysLeuSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerA 450  
QY 6036 GCCAAAACCTCTTAAGTATAGCAACTTCAGCAAGTCTCAGGACACAAATCAATG 6095  
Db 450 laGlnAsnLeuLysLeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnV 470  
QY 6096 TGCAAAATCACAAGCATCTTATACGCCCAATATACACAAACAGAGAGGCAATCATGA 6155  
Db 470 alGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgInnThrGluSerGlnIleMetS 490

QY 6156 GTGAAGTCTCATTCACAATTTGCTACAAAGAGATAAAATACCTAGGATAACACTTACAA 6215  
Db 490 erGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrA 510  
QY 6216 GGGACACGTAAGAACTCTTCAAGGAGAACTACAAACACCTGATCAAGGAATAATAGAGAG 6275  
Db 510 rgAspValLysAspLeuPheLysGluAsnTyrLysProLeuLeu\*\*GluIleLysGluA 530  
QY 6276 ACACAACAATGGAAAAACATTCCTGCTCACAGATAGTAAGAATCAT-----CAAA 6328  
Db 530 sphrAsnLysTyrLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysM 550  
QY 6329 ATGCCATCTGCCCAAGTAATATATAGATTTCAGTCTGCCCTCCATCAAGCTACCATGTA 6388  
Db 550 etAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetT 570  
QY 6389 CTTCCTTCACAGAAATGGAAAAACAATTTAAATTTTCATATATGGAACCAAAAAAGAGCC 6448  
Db 570 hrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAla 589  
QY 6449 CACAGAGCCAAAGACAATCTTAAAGCAAAAGAACTGGAGTATCATGCTACCTGAC 6508  
Db 590 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp 609  
QY 6509 TTAATACTATPACTATAGGCTACAGTAACCAAACTGCATGGTACTGCTACCAAAACAGA 6568  
Db 610 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 629  
QY 6569 TATATAGACCAATGGACAGACAGACCTCAGAAATTTACT- GCAATCTACATCCAT 6627  
Db 630 AspIleAspGlnTrpAsnArgThrGluProSerGluile\*\*\*ProHisIleTyrAsn\*\* 649  
QY 6628 CTGATCTTTTACAAAACCTGCAAAAACAAGCAATGAAAAAGATTCCCTTATTATAAAA 6687  
Db 650 LeuIlePheAspLysPro\*\*\*LysAsnLys\*\*TrpGlyLysAspSerLeuPheAsnLys 669  
QY 6688 TGGTGTGGAAAACTGGCTAGCCATATGCAGAAAGCTGAAACTGGATCCCTTCCTTACA 6747  
Db 670 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 689  
QY 6748 CCTATACAAAAGTTAACTCAAGTGAATTAAGACTTAATATATAGACATATAACCATTA 6807  
Db 690 ProTyrThrLysIleAsnSerArgTyrPheLysAspLeuAsnValArgProLysThrIle 709  
QY 6808 AAAACCCA-CAAGAAAACTTAGCAATACCATTCAGGATATGACATGGCAAGACTTC 6866  
Db 710 LysThrLeuGluGluAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe 729  
QY 6867 ATGACTAAACACACAAAGCAATGGCAACAAAGCCAAATAATAGCAAGTGGGATCTGATT 6926  
Db 730 Met\*\*LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle 749  
QY 6927 AAACTATAGACTCTTCACAGCAAAAACAAACTGCTCATCAGAGTGAAGCAACCTAC 6986  
Db 750 LysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnProTh 769  
QY 6987 AGAATGGGAGAAAAATTTTGCATCTATCGATTCGACAAAGGCTTAATATCCAGATCTTA 7046  
Db 769 r\*\*\*TrpGluLysIlePheAla\*\*TyrSerSerAspLysGlyLeuIleSerArgIleTy 789  
QY 7047 CGAAGAACTTAAACAATTTACAGAAAAA-----ACAACCCCGTCAAAATATGGGCAAA 7102  
Db 789 r\*\*\*GluLeuLysGlnIleTyrLysLysLys\*\*\*AsnAsnProIleLysIleTrpAlaL 809  
QY 7103 GGATATCAGCAGACACTCTCAAAAGAAAGCAATTTATGACAGCCCAACAAACATATGAAAA 7162  
Db 809 sasPmetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLys\*\*\*HisMetLysL 829  
QY 7163 AACCTCATCATCATCTGCTGTAGAGAAATGCAAAACCAACACAGTACCATCATCT 7222  
Db 829 sCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLe 849  
QY 7223 CATGCTAGTTAGAATGGTGTACTATAAAAGTCAGGAAAAACAACAAATGCTGGAGAGATG 7282



```
Db      849  uThrProValhrgMetAlailelleLysLysSerGlyAsnAsnArgCysTrpArgGlyCy 869
QY      7283  TGGAGAAATAGAACACATTTTCCACTGTTGGTGGGAATCTAAATAGTTACCACTTGG 7342
Db      869  sGlyGluileGlyThrLeuLeuHisCysTrpAspCysLysLeuValGlnProLeuTr 889
QY      7343  GAAGACAGTGTGGAGATTCCTTAAGGATCTAGAACACAGAAATATCATTTGACCCAGCAAT 7402
Db      889  pLysSerValTrpArgPheLeuArgAspLeuGluLeuGluileProPheAspProAlaile 909
QY      7403  CCATTACTGAGTATATACCCAAAGAGATATATAATCATCTCTATTATATAAGACACATGCAC 7462
Db      909  eProLeuLeuGlyileTyProLysAspTyLysSerCysCysTyLysAspThrCysTh 929
QY      7463  ACATATGTTTATTGACGACTCATCACATAGCAAGACTTGGAAACCAACCCAAATGCC 7522
Db      929  rArgMetPheileAlaLeuPheThrileAlaLysThrTrpAsnGlnProLysCysTr 949
QY      7523  ATCAGTGTAGACTGGATAAGAAACATATGACATATACACCATGAATACTATGCAGC 7582
Db      949  oThrMetileAspTrpLysLysMetTrpHisileTyThrMetGluTyTrAlaAl 969
QY      7583  CATAAAG-GATGAGTTCATCTCTTTCGACAGATATGATGAGTGAAGTGAACCATCAT 7641
Db      969  aileLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu***ileil 989
QY      7642  TCTCAGCAAACTAACACAGAACAAACCAACCAACCATGTTCTCACTTGTAAAGTGG 7701
Db      989  eLeuSerLysLeuSerGln**GlnLysThrLysHisArg***PheSerLeuileGlycl 1009
QY      7702  GAGT 7705
Db      1009  yAsn 1010
```

RESULT 3

```
US-10-000-256A-153
; Sequence 153, Application US/10000256A
; Publication No. US20030039983A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0259
; CURRENT APPLICATION NUMBER: US/10/000,256A
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/244,782
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 153
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-000-256A-153
```

Alignment Scores:

```
Pred. No.: 1,95e-297 Length: 956
Score: 3770.50 Matches: 783
Percent Similarity: 72.91% Conservative: 65
Best Local Similarity: 67.33% Mismatches: 104
Query Match: 7.21% Indels: 218
DB: 9 Gaps: 7
```

US-10-083-853B-2 (1-29921) x US-10-000-256A-153 (1-956)

```
QY      4244  ATGAGGCAAAATAAAGATGTTCTTTGAAACCAATGAGAACAAAGACACAATGTACCAG 4303
Db      1  MetLysAlaGluileLysValPhePheGluThrAsnGluAsnLysAspThrThrTyrgln 20
QY      4304  AATCTCTGGGCATATTAAAGACAGTGTGTAGAGGGAAATTTATAGCACTAGATGCCTAC 4363
```

```
Db      21  AsnLeuTrpAspThrPheLysAlaValCysArgGlyLysPheileAlaLeuAsnAlaHis 40
QY      4364  AAGAGAAAGCAGGAATATCTAAATAGACACCTTTAAACATCATCAATTTAAAGAACTAGAG 4423
Db      41  LysArgLysGlnGluArgSerLysileAspThrLeuThrSerGlnLeuLysGluLeuGlu 60
QY      4424  AAGAAAGAGCAACAAATTTCAAAGCTAGCAGAACAGACAGAAATAACTAAGATCATCAGACA 4483
Db      61  LysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGlnGluileThrLysileArgAla 80
QY      4484  GAACCTGAGGAGATAGACACACAAAAGCCCTTCAAAATAAATCAATGAATCATCAGAGCTG 4543
Db      81  GluLeuLysGluileGlnThrGlnLysThrLeuGlnLys-ileAsnGlnLysSerArgSerTr 100
QY      4544  GTTTTTTGAAGAGATCAGCAAAATAGACCA----CTAGACAGACTTAATAAACAAGAAAG 4599
Db      100  pPhePheGluArgileAsnLysileAspSerLeuAlaArgLeuileLysLysLysAr 120
QY      4600  AGAGAAAGATCAAGAGATGCAATATAAAGGGATATCAACACCGATCCAC 4659
Db      120  gLulysAsnGlnileAspThrileLysAsnAspLysGlyAspilleThrThrAspProTh 140
QY      4660  AGAAATACAAACTATATATCAGAGAAATATATAACACCTCTATGCAAAATAAATACTAGAAA 4719
Db      140  rGluileGlnThrThrileArgGluTyTrLysHisLeuTyAlaAsnLysLeuGluAs 160
QY      4720  TCTAGAAAGATGGATAATTCCTCGACACATATGTAGCCTGTATGGACCTTGGGGGACA 4779
Db      160  nLeuGluGluMetAspLysPheLeuAsp----- 169
QY      4780  GAACAAAGGGGGTGAATGCAGAAATAAAGACAAAGAGAGATGTATTGGAAGTA 4839
Db      169  ----- 169
QY      4840  GGGGTACAGGGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCACCTCTG 4899
Db      169  ----- 169
QY      4900  TATTTATTAGCAAAAGAGATAGCGAGAGGGTGTGGTGAAGAGAGGTTCAGCTCTTAGG 4959
Db      169  ----- 169
QY      4960  TCCAGAGTAGCCCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCCACTAGA 5019
Db      169  ----- 169
QY      5020  TAACCTCAAGGAGCCAGTGCAGGGAGTGTGGCCCTCAGCAAAACCTTCTAGGGCAGGCA 5079
Db      169  ----- 169
QY      5080  CAGAAAGTAAGTTTGCCACATTTCTGTATTACAGATAAACAGTTTGTCTTTGATCAAGTA 5139
Db      169  ----- 169
QY      5140  GCCTCCAGTGAATGCTGAGTTGGTGCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACAC 5199
Db      170  -----Th 170
QY      5200  ATACACCTCTCAAGACTAAACAGGAAGAGTCAATCCCTGATATATACAGTAAACAAG 5259
Db      170  rTyThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProileThrGl 190
QY      5260  TTCTAAATTCAGACGATTAATGTATAGCTTACCACCAAAAAAGTCCAGGACCAAGCGG 5319
Db      190  yAlaGluileValAlaileileAsnSerLeuProThrLysLysSerProGlyProAspGl 210
QY      5320  ATTCACAGCCAAATTTCTACAGAGGTACAAAAGAGAGAGTGTGTACTATTCTTCTGAAACT 5379
Db      210  yPheThrAlaGluPheTyrglnSerTrpAlaGlu----- 221
QY      5380  ATTCACAAATAATAGAAAATGGGAATCCTCCCTAACTCATTTTACGAGGCCAGCATCATCC 5439
```

Db	221	-----	221
Qy	5440	TGATACCAAAACCTAGCAGTCACACAACAAAGAGAAATTTTCAGGCCATATCCCTGA	5499
Db	222	---Thr-GlnPro-----LysLysGluAsnPheArgProIleSerLeuM	235
Qy	5500	TGAACATTGATGTGAAATCCTCAATAAATCTGCGAAACCAACCAATCCAGCAGCACATCA	5559
Db	235	etAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaLysArgIleGlnHisIleL	255
Qy	5560	AAAAGCTTATCTACCATGATCAAGTTGGCGTCATCCCTGGGATCGAAGCTGTTCAAAA	5619
Db	255	ysLysLeuIleHisHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnI	275
Qy	5620	TATCAAAATCAATAATGTAGGCCATCACATAAAGACAAACCAATGACAAAACACACATGA	5679
Db	275	leArgLysSerIleAsnValThrGlnHisIleAsnArgAlaLysAspLysAsnHisMetI	295
Qy	5680	TTATCTCAATAGATGCAGAAAGCCCTTGTCAAAATTCACAGCCCTTCATGCTAAAAA	5739
Db	295	leIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysT	315
Qy	5740	TTCTCAGTAAACTAGGTATCGATGGAATGTATCTCAAAATTAATAGAGCTATTATAC-A	5798
Db	315	hrLeuAsnLysLeuGlyIleAspGlyThrPheLysIleIleArgAlaIleIleIleIleI	335
Qy	5799	AACCCACAGCCAAATATCATCTGAATGGCAAAACCTGGAAGCATTCCTTTGAGAAGCTG	5858
Db	335	snProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrG	355
Qy	5859	GCACAAGCAGAGGATGCCCTCTCACACTCCTATTTCAGATACTACTTGAAGTCTTGG	5918
Db	355	lyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluValLeuA	375
Qy	5919	CCAGGCAATCAGGCAATAGAAAGAAATAAAGGCTATTCAATTAAGAAGCAGCAAGTCA	5978
Db	375	laArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyIleGluValL	395
Qy	5979	TATTGTCTCTGTTTCAGATGACATGTTGTGTATATTAGAAAACCCCATCTCTCAGGCC	6038
Db	395	ysLeuSerLeuPheAlaAspAsnMetIleValTyLeuGluAsnProIleValSerAlaG	415
Qy	6039	AAAACTCTTAAGCTGATAGCAACTTCACCAAGTCTCAGGACACAAAATCAATGTGC	6098
Db	415	InAsnLeuLeuLysLeuIleSerAsnPheSerLysValSerGlyTyLysIleAsnValG	435
Qy	6099	AAAAATCACAAGCATCTTATACGCCAATAATACACAAACAGACAGCCAAATCATGAGTG	6158
Db	435	InLysSerGlnAlaPheLeuTyThrAsnAsnArgGlnThrGluSerGlnIleMetSerG	455
Qy	6159	AACCTCTCATTCACAAATGCTACAAAGAGATAAATAACCTAGGAATACAACTTACAAGGG	6218
Db	455	InLeuProPheThrIleAlaSerLysArgIleLysTyLeuGlyIleGlnLeuThrArgA	475
Qy	6219	ACAGTGAAGAACTCTCAAGGAGAACTACAAACCACTGATCAAGGAATAAGAGAGACA	6278
Db	475	spValLysAspLeuPheLysGluAsnTyLysProLeuLeuLysGluIleLysGluAspT	495
Qy	6279	CAAAACAATGGAACAAATCCATGCTCACAGATAGTAAGATCAT-----GAAATG	6331
Db	495	hrAsnLysTrpLysAsnIleProCysSerGlyGluGlyArgIleAsnIleValLysMetA	515
Qy	6332	CCATACCTGCCAAAGTAAATATAGATTGCTAGCTACGCCCATCAAGCTACCATTGACTT	6391
Db	515	lalleLeuProLys-----	519
Qy	6392	TCITTCACAGATTCGAAAAACAACTTTAAATTTTCATATGGAACCAAAAAAGACCCAC	6451
Db	520	-----GluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaHis	536
Qy	6452	AGACCCAGACAATCTTAAGCAAAAAAGACAAACCTGGAGGTATCATGCTACCTGACTTA	6511
Db	537	IleAlaLysSerIleLeuAsnGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPhe	556
Qy	6512	AACTATACATATAGGCTACAGTAACCAAACTGCATGGTACTGCTACCAAAACAGATAT	6571
Db	557	LysLeuTyTyLysAlaThrValThrLysThrAlaTrpTyTrpTyGlnAsnArgAsp	576
Qy	6572	ATAGACCAATGGAACAGACAGACGCTCAGAAATTCACCT-GCAATCTACATCCATCTG	6630
Db	577	IleAspGlnTrpAsnArgThrGluProSerGluIleThrGlnHisIleIleSerTyLeu	596
Qy	6631	ATCTTTTACAAACCTGACAAAACAACAAGCAATGGAAGGATTCCTTATTAAATAATGG	6690
Db	597	IlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrp	616
Qy	6691	TGTTGGAAAAAAGCTAGGCATATGAGAAAGCTGAAACTGGATCCCTTCTTACACCT	6750
Db	617	CysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrPro	636
Qy	6751	TATCAAAAGTTACTCAAGTGAATTAAGACATTAATAATAGACATATAAACCATAAAA	6810
Db	637	TyrThrLysMetAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLys	656
Qy	6811	ACCCA-GAAGAAAACCTAGGCAATACCATTCAGCATATGGACATGGGCAAGACTTCATG	6869
Db	657	ThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPheMet	676
Qy	6870	ACTAAAAACACAAAGCAATGGCAACAAAGCCAAAATAGACAAAGTGGATCTGATTAAA	6929
Db	677	SerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuValLys	696
Qy	6930	CTATAGACCTCTCCACAGCAAAAANAACCTGTCATCAGAGTGAACAGCAACCTACAGA	6989
Db	697	LeuLysSerPheCysThrAla-LysGluThrIleArgValAsnArgGlnProThrLy	716
Qy	6990	ATGGCAGAAAATTTTCAATCTATCGATCTGACAAAGGCTAATATCAGAGATCTACGA	7049
Db	716	stfGluLysIlePheAlaThrTySerSerAspLysGlyLeuIleSerArgIleTyAs	736
Qy	7050	AGAACTTAAACAAATTTACAGAAAAA-----AACACCCCGCTCAAAATATGGCAAGGA	7105
Db	736	nGluLeuLysGlnIleTyLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAs	756
Qy	7106	TATCAGCAGACACTTCTCAAAAGAAGACATTTATGACGCCAACAAACATATGAAAAAAC	7165
Db	756	pMetAsnArgHisPheSerLysGluAspIleTyAlaAlaLysLysHisMetLysLysCy	776
Qy	7166	CTCATCATCATGCTGCTTAGAGAAATGCAAAACAAACACACAGTGCATACCATCTCAT	7225
Db	776	sSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyHisLeuTh	796
Qy	7226	GCTAGTTAGATGTGATCATAAAAGTCAGGAAACAAACAAATCTGCTGGAGAGATGTGG	7285
Db	796	rProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGl	816
Qy	7286	AGAAATAGGACACACTTTCCACTCTGCTGGTGGGAATGAATTAAGTTCACCATTTGTGAA	7345
Db	816	yGluThrGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuAlaGlnProLeuTrpLy	836
Qy	7346	GACAGCTGGAGATTCCTTAAGGATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCC	7405
Db	836	sSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIlePr	856
Qy	7406	ATTACTGAGTATATACCAAGGAATATAAATCTTCTATTATTAAGACACATGCACACA	7465
Db	856	oLeuLeuGlyIleTyProLysAspTyLysSerCysCysTyLysAspThrCysThrAr	876
Qy	7466	TATCTTTATGACACACTGATCACAATAGCAAGACTTGGAAACCAACCAATGTCATC	7525
Db	876	gMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProTh	896
Qy	7526	AGTCATAGACTGGATAAAGAAACATGGCACATATACACCATGAATACTATGAGCCAT	7585
Db	896	rIleIleAspTrpIleLysLysMetTrpHisIleTyThrMetGluTyTyAlaAlaIle	916



```
QY 5581 AAGTGGCGCATCCCTGGGATGCAAGCGCTGGTTCAAAATATGCAAAATCAATAATGTAG 5640
D 5582 AAGTGGCGCATCCCTGGGATGCAAGCGCTGGTTCAAAATATGCAAAATCAATAATGTAG 5640
D 352 InValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIleAsnValI 372
QY 5641 GCATCATCAATAACAGAACCAATGACAAAACACATGATATCTCAATAGATGACGAAA 5700
D 372 leGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAspAlaGluL 392
QY 5701 AGGCTTTGTGCAAAATCAACAGCCCTTCATGCTTAAATTCCTCACTAACTAGGTATCG 5760
D 392 ysaIaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleA 412
QY 5761 ATGGAATGTATCTCAAAATATAAGAGCTATTTATAC-AAACCCACACCAATATCATAC 5819
D 412 spGlyThrTyPheLysIleIleArgAlaIleTyAspLysProThrAlaAsnIleIleL 432
QY 5820 TGAATGGGCAAAACACTGGAGCATCTCCCTTTGAGAACTGGCAAGCAAGGATGCGCTC 5879
D 432 euAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGlyCysProL 452
QY 5880 TCTCACCACCTCTATTCAAGATACTATTGGAAGTCTCGCCAGGCAATCAGGCAATAGA 5939
D 452 euserProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgGlnGluL 472
QY 5940 AAGCAATTAAGGCTATTCAATAGAAAAGAGGAGGATCATATTGCTCTCTTTGCGATG 5999
D 472 ysgLulLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspA 492
QY 6000 ACATGTTGTATATTAGAAAACCCCATCTCTCAGGCCAAAACCTCTTAAGCTGATAA 6059
D 492 spMetIleValTyLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLeuIleS 512
QY 6060 GCACTTCAGCAAGTCTCAGGACACAAATCAATGTGCAAAATCAACAGCATCTCTAT 6119
D 512 erAsnPheSerLysValSerGlyTyLeuLysIleAsnValGlnLysSerGlnAlaPheLeu 532
QY 6120 ACGCCAATAATAGACAAACAGAGCCAAATCATAGTGAACCTCTCATTCACAAATGCTA 6179
D 532 yThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIleAlaS 552
QY 6180 CAAGAGCAATAAATCACTAGGATACAACTTACAGGGACACGTAGGAATCTTCAAGG 6239
D 552 erLysArgIleLysTyLeuGlyIleGlnLeuThrArgAspValLysAspLeuPheLysG 572
QY 6240 AGAACAACAACCACTCATCAAGSAAATAGAGAGGACACAAACAATGAAAACATTC 6299
D 572 luAsnTyLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLysAsnIleP 592
QY 6300 CATGCTCACAGATAGTAAGATCAT-----GAAAATGCCATCTGCCCCAAAGTAAAT 6352
D 592 rocYsSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLysValIle 612
QY 6353 ATAGATTCAGTGTACCCCATCAAGCTACCATGACTTCTTCCACAGAAATGGAAGAAA 6412
D 612 yArgPheAsnAlaIleProIleLysLeuProMetThrPhePheThrGluLeuGluLys 632
QY 6413 CAACTTTAAATTCATATGAACCAAAAAGAGCCACAGAGCCAGACAACTTTAAGC 6472
D 632 hrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAlaLysSerIleLeuSer 651
QY 6473 AAAAGAACAAAGCTGGAGGTATCATGCTACCTGACTTAAAACTACTATAAGGCTTACA 6532
D 652 GlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyTyLysAlaThr 671
QY 6533 GTAACCAAAACTGAGGTACTGGTACCACAAACAGATATATAGACCAATGGACAGACA 6592
D 672 ValThrLysThrAlaTrpTyTrpTyTrpGlnAsnArgAspIleAspGlnTrpAsnArgThr 691
QY 6593 GAGACCTCAGAAATTAC-ACTGCAATCTACATCCATCTGATCTTTGACAAACCTGACAAA 6651
D 692 GluProSerGluIleMetProHisIleTyAsnTyLeuIlePheAspLysProGluLys 711
```

```
QY 6652 AACAGCAATGGAAGGATCCCTATTATAAATGTTGGTGGAAAACTGGCTAGCC 6711
D 712 AsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAla 731
QY 6712 ATATGCAAGAAAGCTGAACACTGGATCCCTTCTTACACCTTATACAAAAGTTAACTACA 6771
D 732 IleCysArgLysLeuLysLeuAspProPheLeuThrProTyTrpLysIleAsnSerArg 751
QY 6772 TGAATTAAGACACTTAATATAACACATAAAACCAATAAAACCCA-GAAGAAAAACCTAGGC 6830
D 752 TrpLleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluGluAsnLeuGly 771
QY 6831 AATACCATTCAGGATATGACATGGCAAAAGACTTTCATGACTAAAACACCAAAAGCAATG 6890
D 772 IleThrIleGlnAspIleGlyValGlyLysAspPheMetSerLysThrProLysAlaMet 791
QY 6891 GCAACAAAAGCCAAATAGACAAGTGGGATCTCTATTAAACTATAGAGCTTCTGCACAGCA 6950
D 792 AlaThrLysAlaLysIleAspLysTrpAspLysLeuIleLysLeuLysSerPheCysThrAla 811
QY 6951 AAAAAAACTGTCTCAGAGTGAAACAGCAACCTACAGAATGGAGAAAAATTTTGGCAAT 7010
D 812 -LysGluThrThrIleArgValAsnArgGlnProThrThrTrpGluLysIlePheAlaTh 831
QY 7011 CTATGATCTGCAGAAAGGCTTAATATCCAGAGATCTACGAAGAATCTTAAACAAAATTTACAA 7070
D 831 rTySerSerAspLysGlyLeuIleSerArgIleTyAsnGluLeuLysGlnIleTyL 851
QY 7071 GAAAAA-----AACACCCCTCAAAATATGGGCAAGGATATGAGCAGACACTTCTCAAA 7126
D 851 sLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArgHisPheSerL 871
QY 7127 AGAAGACATTATGCGCCCAACACATATGAAAAAACCTCATCATCTGTCGTTAG 7186
D 871 sGluAspIleTyAlaAlaLysLysHisMetLysLysCysSerSerLeuAlaIleAr 891
QY 7187 AGAAATGCAAAACAAACACAGTACATACCATCTCATCTAGTTAGATAGTGTATCAC 7246
D 891 gGluMetGlnIleLysThrThrMetArgTyHisLeuThrProValArgMetAlaIleIl 911
QY 7247 TAAAAAGTCAGGAACACAC 7265
D 911 eLysLysSerGlyAsnAsn 917
RESULT 5
US-10-082-830-235
; Sequence 235, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082, 830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243, 802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 235
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-235
Alignment Scores: 2.58e-278 Length: 940
Pred. No.: 3535.50 Matches: 747
Score: 74.43%
Percent Similarity: 74.43%
Conservative: 68
```

Best Local Similarity: 68.22% Mismatches: 119  
Query Match: 6.76% Indels: 167  
DB: 9 Gaps: 7

US-10-083-853B-2 (1-29921) x US-10-082-830-235 (1-940)

```
QY 4016 GAATATACACTCTTCTCAGCATCATACACTTAAATTTTAAATTTGACCATGTAATTTTA 4075
||||| ||||||| ||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 GluTyrThrSerPheSerAlaLeuHisAsnThrTyrSerLysIleAspHisIleValGly 20
QY 4076 AGTAAACACTCTCAGCAAAATGCAAAAGACAGAAATCCCTAACAAACAGTCTCTCAGAC 4135
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 21 SerLysAlaLeuSerLysCysLysArgThrGluMetIleThrAsnCysLeuSerAsp 40
QY 4136 TACAGTGAATCTATTAGAACTCAGAAATTAAGAACTCACAATACACAACTACA 4195
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 41 HisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnCysSerThrThr 60
QY 4196 TGGAACTGAACAACCTGCTCTGAATGACTACTGGTAAATPACAAAATGAAGCAAAA 4255
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 TrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrCysValHisAsnLysMetLysAlaGlu 80
QY 4256 ATAAAGATGCTCTTGAACCAATGAGAACAGACACAATGTACCAGAACTCTCGGGC 4315
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 81 IleLysMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAsp 100
QY 4316 ATATTAAAGCAGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAAGAGAAAGCAG 4375
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 101 ThrPheLysAlaValCysArgLysAsnPheIleAlaLeuAsnValHisLysArgLysGln 120
QY 4376 GAAATATCTAAATAGACACCTTAACATCAATTAAGAACTAGAGAAGAAAGAGCAA 4435
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 GluArgSerLysIleAspThrLeuIleSerGlnLeuLysGluLeuGluLysGlnGluGln 140
QY 4436 ACAAAATCAAAAGCTAGCAGAGACAGAAATACTAGATCAGACAGAACTGGAAGAG 4495
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 141 ThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaGluValLysGlu 160
QY 4496 ATAGACACAAAAAGCCCTCAAAATAATCAATCAATCCAGAGCTGTTTTTGAAGA 4555
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 161 IleGluThrGlnLysThr-PheLysArgIleAsnGluSerArgAsnTrpPheGluAr 180
QY 4556 GATCAGCAAAAT---AGACCACTAGACAGACTAATAAGAGAAAGAGAGAAATCA 4611
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 180 GileSerLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnG 200
QY 4612 AAGAGATCAATAAAAA---AATGATAAAGGGGATATCACCACCGATCCACAGAAATACA 4668
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 200 nileaspAlaIleAsnThrHisAspLysGlyAspIleThrAspProThrGluIleG 220
QY 4669 AACTATTATCAGAGAAATATTATAACACCTCTATGCAAAATAACTAGAAAAATCTAGAAGA 4728
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 220 nThrThrIleArgGluTyrTyrLysHisPheTyrAlaAsnLysLeuGluAsnLeuGluG 240
QY 4729 AATGATAAAATTCCTGGACACATATGTAGCCCTGTATGACCTTGGGGGACAGACAAAAG 4788
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 240 uMetAspLysPheLeuAsp----- 246
QY 4789 GGGGTGAATGCAGAAATAAAGACAAAGACAAAGAGATATGTTTGAAGTAGGGTCAGG 4848
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4849 GGGCAACTTGGCTCTAATGGACAGGGCCCTGAGCTTTACACCACCCCTCTGTATTATTA 4908
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4909 GCGAAAGAGATAGCGAGGGTGATTTGGAAGAAGAGGTCAGCTGTAGTCCAGAGTA 4968
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
QY 4969 GGCCTGCAAGACTGCTTCTCAACAANTAGCTCTAGATGTCCAGTAGATAACCTCAA 5028
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 ----- 246
```

```
QY 5029 GGAGCCAGTGCAGGAGGAGTGGTGGCCCTCAGCAAACTTTCTAGGCGAGCACAGAAATAA 5088
246 ----- 246
QY 5089 GTTTGCCCCACATCTGTGATTACAGATAAACAGTTTGTGTTGATCAAGTAGTCCCTCCAGT 5148
246 ----- 246
QY 5149 GGAATGCTGAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACATACACCT 5208
247 ----- 247
QY 5209 CTCAGACTAAACAGGAAGTCAAAATCCCTGTAATATACCAAGTAAACAGTTCTTAAAT 5268
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 250 uProArgLeuAsnGlnGluAlaGluSerLeuAsnArgProIleThrAspSerGluI 270
QY 5269 TGAAGCAGTAATGATGATCCCTACCAACCAAAAGTCCAGGACAGGATTCACAGC 5328
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 270 eAlaAlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrPr 290
QY 5329 CAAATTCACAGAGGTACAAAGAGAGCTGTACTATTCTCTGAACTATTCCAAAAA 5388
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 290 oLysPheTyrGlnArgTyrLysGluIleuValProPheLeuLeuLysLeuPheGlnSe 310
QY 5389 AATA---GAAATGGGAATCCTCCCTAACTCATTTTACGAGCGCCAGCATCTCTGTATAC 5445
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 rIleThrLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaAsnIleIleLeuIleL 330
QY 5446 CAAACCTAGCAGTGACACACACAAAAGAGAAATTTTCAGGCCCATCTCCCTGATGAACA 5505
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 330 euLysProGlyArgAspThrThrLysLysArgGluPheArgProIleSerMetMetIleI 350
QY 5506 TTGATGTCAAAATCCTCAATAAAATACTGGCAAACTCAAAATCCAGCAGCACATCAAAAAGC 5565
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 350 leAspAlaLysIleLeuSerLysIleLeuAlaAsnGlnIleGlnHisLeuIleLysL 370
QY 5566 TTATCTACCATGATCAAGTTGGCGTCACTCCCTGGGATGCAAGCGTGGTTCAAAATATGCA 5625
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 370 euIleHisAspGlnValGlyPheIleProGlyMetLysGlyTrpPheAsnIleArgL 390
QY 5626 AATCAATAAATGTAGGCCATCACATAAACAACAACCAATGACAAAACACACATGATTATCT 5685
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 ysSerIleLysValIleHisIleAsnArgThrLysAspLysAsnHisMetIleIleI 410
QY 5686 CAATAGATGCAAGAAAGCCCTTTGTCAAAATTCACAGCCCTTCATGTCAAAATTCCTCA 5745
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 410 erIleHisAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrValA 430
QY 5746 GTAACTAGGTATCGATGGAATGTATCTCAAAATAAATAAGAGCTATTATTATAC-AAACCCA 5804
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 430 snLysLeuValIleAspGlyThrTyrLeuLysIleIleArgAlaIleTyrAspLysProt 450
QY 5805 CAGCCATATCATACTGNAATGGGCAAAACTGGAGCATTCCTCTTGAGAACTGGCACAA 5864
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 450 hrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuArgThrGlyIleA 470
QY 5865 GACAAGGATGCCCTCTCTCACCACTCTTATTCAAGATACTATTGGAAGTTCTGGCCAGGG 5924
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 470 rGlnGlyCysProLeuSerProLeuPheAsnIleValLeuGluValLeuAlaArgA 490
QY 5925 CAATCAGGCAATAGAGAAATAAAGGATTTCAAAATAGAAGAGAGAGGAGTCAATTGT 5984
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 490 laIleArgGlnLysGluIleLysGlyIleGlnLeuGlyLysGluLysValLysLeuS 510
QY 5985 CTCCTGTTGAGATGACATGTTGTATATTTAGAAAACCCCATCGTCTCAGGCCCAAAAC 6044
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 510 erLeuPheAlaAspMetIleLeuTyrLeuGluAsnProIleValSerAlaGlnAsnL 530
QY 6045 TCCTTAAGCTCATAGAACCTTCAGAAAAGTCTCAGGACACAAAATCAATGTGAAAAAT 6104
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 530 euLeuLysLeuMetSerSerPheSerLysValSerGlyTyrLysIleAsnValGlnLysS 550
```

QY 6105 CACAAGATTCTTTATAGCCCAATAATAGACAAACAGAGAGCCAAATCATGAGTGAATCT 6164  
Db 6106 CACAAGATTCTTTATAGCCCAATAATAGACAAACAGAGAGCCAAATCATGAGTGAATCT 6165  
QY 6165 CATTACAAATGCTACAAAGAGAAATAAATACCTAGGAATACAACTTACAGAGGACAGCT 6224  
Db 6166 CATTACAAATGCTACAAAGAGAAATAAATACCTAGGAATACAACTTACAGAGGACAGCT 6225  
QY 6225 AGGAATCTTCAAGAGAGAACTCAACACCACTGATCAAGGAATAAAGAGAGACACAAACA 6284  
Db 6285 ysaPleuPheLysGluAsnTyLysProLeuLeuAsnLysLysGluAspThrAsnL 609  
QY 6285 AATGGAATAACATCCATGCTACAGATAGTAAGAAATCAT-----GAAATGCCATAC 6337  
Db 6337 ystrLysAsnLysProCysSerTripleGlyArgLysLeuValLysMetAlaIle 629  
QY 6338 TGCCCAAGCTAAATATAGATTCAGTACCTACCCCATCAAGCTACCATTCACCTTTCTCA 6397  
Db 6397 etProLysValIleTyArgPheAsnAlaIleProLysLeuProMetThrPhePhe 649  
QY 6398 CAGAATTTGAAAAACAACCTTTAAATTTTCATATGGAACCAAAAAAGAGCCACAGAGCC 6457  
Db 6458 hrcLyuLeuLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgLysAla 668  
QY 6458 AGACAATCTTAAGCAAAAAAGAAAGCTGGAGTATCATGCTACCTTAAACTA 6517  
Db 6517 LysThrIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeu 688  
QY 6518 TACTATAAGCTACAGTAACCAAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6577  
Db 6577 TyrTyLysAlaThrValThrLysThrAlaTrpTyTrpGlnAsnArgAspIleAsp 708  
QY 6578 CAATGGAACACAGACAGCTCAGAAATTTAC-ACTGCAATCTACATCCATCGACTTT 6636  
Db 709 GlnTrpAsnArgIleGluProLeuGluLeuProHisIleTyTrpAsnHisLeuIlePhe 728  
QY 6637 GACAAACCTGACAAAAACAGCAATGGAAGGATTCCTATTTTAAATGAATGGTGTGG 6696  
Db 729 AspLysProAspLysAsnLysLeuTrpGlyLysAspSerLeuPheAsnLysTrpCysTrp 748  
QY 6697 AAAAAGCTGCTAGCCATATGCAAGAGCTGAAAGCTGATCCCTTCCTTACACCTTATACA 6756  
Db 749 GluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnLeuPheLeuThrProTyTrp 768  
QY 6757 AAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCCA- 6815  
Db 769 LysIleAspSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysIleLeu 788  
QY 6816 GAAGAAACCTAGGCAATCCATTCAGGATATGGACATGGGCAAGACTTCATGACTAA 6875  
Db 789 GluLysAsnLeuGlyAsnThrIleGlnAspIleGlyValGlyLysAspPheMetThrLys 808  
QY 6876 ACACCAAAAGCAATGCAACAAAGCCAAATAGACAGTGGATCTGATTAACTATAG 6935  
Db 809 ThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspIleLysLeuLys 828  
QY 6936 AGCTTCTGCACACAAAAAAGCTGCTACAGTGAACCAACCACTACAGAAATGGGA 6995  
Db 829 SerPheCysThrAla-LysGluThrThrIleIleValAsnArgGlnProThrGluTrpG 848  
QY 6996 GAAATTTTGCATATTCATGCTGACAAAGGCTAATATCCAGAGATCTACAGAAACT 7055  
Db 848 uLysIlePheLysIleTyProSerAspLysGlyLeuIleSerArgIleTyLysGluLe 868  
QY 7056 TAAACAAATTTACAAAAA-----AACAAACCCCTCAAAATATGGCAAGGATATGAG 7111  
Db 868 uLysGlnIleTyLysLysSerAsnAsnProIleLysAsnTrpAlaLysAspMetAs 888  
QY 7112 CAGACACTTCTCAAAAAGACATTTATGAGCAACCAACATATGAAAAAACCCTCATC 7171  
Db 888 nArgHisPheSerLysGluAspIleTyAlaValAsnArgHisMeLysThrCysSer 908  
QY 7172 ATCATTTGGTGTAGAGAAATGCAAAAAACCAACCAAGTACATACCATCTCATGCTAGT 7231

Db 908 rLeuLeuAlaIleArgGluMetGlnIleLysThrMetArgTyHisPheThrProva 928  
QY 7232 TAGAATGGTGATCACTAAAAAGTCAGGAACACAA 7268  
Db 928 lArgMetAlaSerIleLysLysSerGlyAsnAsnArg 940  
RESULT 6  
US-10-025-201-3  
; Sequence 3, Application US/10025201  
; Publication No. US20030003468A1  
; GENERAL INFORMATION:  
; APPLICANT: Crow, Mary K.  
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY  
; FILE REFERENCE: 5983/2H567  
; CURRENT APPLICATION NUMBER: US/10/025,201  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,673  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1275  
; TYPE: PRT.  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U091116  
; DATABASE ENTRY DATE: 1995-02-02  
; RELEVANT RESIDUES: (1)..(1275)  
US-10-025-201-3  
Alignment Scores:  
Pred. No.: 8,79e-171 Length: 1275  
Score: 2215.00 Matches: 408  
Percent Similarity: 95.01% Conservative: 11  
Best Local Similarity: 92.52% Mismatches: 22  
Query Match: 4.23% Indels: 0  
DB: 9 Gaps: 0  
US-10-083-853B-2 (1-29921) x US-10-025-201-3 (1-1275)  
QY 29919 ACTTCTTCCAGAAATGGAAAACTACTTTAAAGTTCATGTGGAAACCAAAAAAGAGCC 29860  
Db 835 ThrPhePheThrGluLeuGluLysThrLeuLysPheIleTrpAsnGlnLysArgAla 854  
QY 29859 CACATCGCAAGTCAATCTTAAGCAAAAGAAAGCAAGTGGAGCATCATGCTACCTGAC 29800  
Db 855 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProTy 874  
QY 29799 TTCAAACTACACTACAGGCTACAGTAACCAAAACCAACATGGTACTGGTACCAAAACAGA 29740  
Db 875 PheLysLeuTyTrpLysAlaThrValThrLysThrAlaTrpTyTrpGlnAsnArg 894  
QY 29739 GATATCAACCAATGGAAACAGACAGAGCCCTCAGAATAATATGCCGATATCTCAACTAT 29680  
Db 895 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyTrpAsn 914  
QY 29679 CTGATCTTGCACAAACCTTGAGAAAAACAAGCAATGGGAAGGATTCCTTATTATAAA 29620  
Db 915 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 934  
QY 29619 TGGTGTGAGAAACTGGGTAGCCCTATGTAGAAAGCTGAAACTGGATCCCTTCCTTACA 29560  
Db 935 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspLeuPheLeuThr 954  
QY 29559 CCTATACAAAATAATTAAGTCAAGATGGAATTAAGACTTAAGTCTTAGACTTAAACCAT 29500  
Db 955 ProTyThrLysIleAsnSerArgTrpIleLysAspLeuAsnValLysProLysThrIle 974  
QY 29499 AATCCCTAGAACAAAAACCCAGGCAATACCATTCAGACATAGGCATGGCAGGACTTC 29440  
Db 975 LysThrLeuGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAspPhe 994



QY 29439 ATGCTAAACACCAAGCAATGGCAACAAAGCCAAATTCAGAAATGGGATTAATT 29380  
Db 995 MetSerLysThrProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuIle 1014  
QY 29379 AAACCTAAAGAGCTTCTGCACACAAAGAAATACCATCAGAGTGAACAGGCAACTTACA 29320  
Db 1015 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 1034  
QY 29319 GAATGGGAGAAATTTTGCACACCTACTCATTTGCAAGGGCTAATATCCAGATCTAC 29260  
Db 1035 ThrTrpGluLysIlePheAlaThrTrpSerSerAspLysGlyLeuIleSerArgIleTrp 1054  
QY 29259 AATGAACCTAAACAAATTTTACAAGAAAAAACAAGACCCCATCAAAAGTGGGTGAAG 29200  
Db 1055 AsnGluLeuLysGlnIleTrpLysLysThrAsnAsnProIleLysLysTrpAlaLys 1074  
QY 29199 GATATGAACAGACACTTCTCAAGAGAGGCATTTATGAGCCCAAAAACACATGAAAAA 29140  
Db 1075 AspMetAsnArgHisPheSerLysGluAspIleTrpAlaAlaLysLysHisMetLysLys 1094  
QY 29139 TACTCATCATCACTGGCCATCAGAGAAATGCAAAATCAAAACCAATGAGATACCATCTC 29080  
Db 1095 CysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTrpHisLeu 1114  
QY 29079 ACACGAGTTAGATGGCAATCAATTAAGAAAGTCAAGAAACACAGGTGCTGGAGAGATGT 29020  
Db 1115 ThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCys 1134  
QY 29019 GGAGAAATAGGAACACTTTTACACTGTTGGGACTGTGACTAGTTCACCATCTGGG 28960  
Db 1135 GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 1154  
QY 28959 AAGTCAGTGTGTGATTCCTCAGGATCTAGAACTAGAAATACCAATTTGACCCGCCATC 28900  
Db 1155 LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 1174  
QY 28899 CCATTACTGGGTATATACCAAGGATTATAATCATGCTGCTATAAAGACACAAAGCACA 28840  
Db 1175 ProLeuLeuGlyIleTrpProGluAspTrpLysSerCysTrpLysAspThrCysThr 1194  
QY 28839 TGTATGTTATAGCAGCACTATTCACATAGCAAGACTTGAACCAACCTAAATGTCCA 28780  
Db 1195 ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 1214  
QY 28779 ACAAGCATAGACTGATTAAGAAAAATGTGGACATATACACCATGGAATCTATCGAGCC 28720  
Db 1215 ThrMetIleAspTrpIleLysLysMetTrpHisIleTrpThrMetGluTrpTyrAlaAla 1234  
QY 28719 ATAAAAATGATGAGTTCGTGCTTCCTTTAGGACATGGATGAAGCTGGAACCATCAT 28660  
Db 1235 IleLysAsnAspGluPheIleSerPheValGlyThrTrpMetLysLeuGluThrIleIle 1254  
QY 28659 CTCAGCAACTATCACAGGACAAAAAACCAACACCCGATGTTCTCACTCATGATGGG 28600  
Db 1255 LeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGlyGly 1274  
QY 28599 AAT 28597  
Db 1275 Asn 1275

RESULT 7

US-10-050-882-80  
; Sequence 80, Application US/10050882  
; Publication No. US20030104400A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 27 Human secreted proteins  
; FILE REFERENCE: P2038P1  
; CURRENT APPLICATION NUMBER: US/10/050,882  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: 09/661,453  
; PRIOR FILING DATE: 2000-09-13  
; PRIOR APPLICATION NUMBER: PCT/US00/06783

; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: 60/125,055  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 156  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 80  
; LENGTH: 1010  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (25)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (104)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (194)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (362)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (525)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (643)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (649)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (656)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (660)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (731)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (770)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (777)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (790)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (800)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (825)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (987)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (996)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (1003)  
; OTHER INFORMATION: xaa>equals any of the naturally occurring L-amino acids  
US-10-050-882-80

Alignment Scores:

Pred. No.: 1.57e-169 Length: 1010  
Score: 2199.00 Matches: 405  
Percent Similarity: 93.42% Conservative: 7  
Best Local Similarity: 91.84% Mismatches: 29  
Query Match: 4.20% Indels: 0  
DB: 9 Gaps: 0



US-10-083-853B-2 (1-29921) x US-10-050-882-80 (1-1010)	
QY 29919	ACTTTCTTCACAGAAATGGAAAAAAGTACTTTAAAGTTTCATGTGGAAACCAAAAAAGAGCC 29860
Db 570	ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 589
QY 29859	CACATGCCCAAGTCAATCCCTAAGCCAAAAGAAAGAAAGCTGGAGGCATCATGCTACCTGAC 29800
Db 590	ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAsp 609
QY 29799	TTCAAACTACATACAGGCTACAGTAACCAAAACAAACATGGTACTGGTACCAAAACAGA 29740
Db 610	PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArg 629
QY 29739	GATATCAACCAATGGAGACAGAGCCCTCAGAAATATATCCGCATATCATCACTAT 29680
Db 630	AspIleAspGlnTrpAsnArgThrGluProSerGluIle***ProHisIleTyrAsn*** 649
QY 29679	CTGATCTTTGACAAACCTGAGAAAAACAAGCAATGGGAAAGATTCCCTATTATAATAA 29620
Db 650	LeuIlePheAspLysPro***LysAsnLys***TrpGlyLysAspSerLeuPheAsnLys 669
QY 29619	TGCTGCTGAGAAACTGGCTAGCCCTATGTAGAAAGCTGAAAGTGGATCCCTTCCTTACA 29560
Db 670	TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 689
QY 29559	CCTTATACAAAAATTAATCAAGATGATTAAGACTTAAATGTTAGACTTAACCACTA 29500
Db 690	ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 709
QY 29499	AAATCCCTAGAAGAAACCCAGCAATACCATTCAGCAGATAGGCTGGCAGGACTTC 29440
Db 710	LysThrLeuGluLysLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPhe 729
QY 29439	ATGCTTAACACCAAAAGCAATGGCAACAAAGCCAAATGACAAATGGGATCTAAT 29380
Db 730	Met***LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle 749
QY 29379	AAACTAAAGAGCTTCGCAGACAAAGAAAGTACCATCAGAGTGAACAGGCAACTTACA 29320
Db 750	LysLeuLysSerPheCysThrAlaLysGluThrIleArgValAsnArgGlnProThr 769
QY 29319	GAATGGAGAAAAATTTTGCACCTACTCATTTGACAAAGGGCTAATATCCAGAATCTAC 29260
Db 770	***TrpGluLysIlePheAla***TyrSerSerAspLysGlyLeuIleSerArgIleTyr 789
QY 29259	AATGAATCAACAAATTTACAGAAGAAAAACAAAGAACCCCATCAAAAGTGGTGAAG 29200
Db 790	***GluLeuLysGlnIleTyrLysLysLys***AsnAsnProIleLysLysTrpAlaLys 809
QY 29199	GATATCAACAGACACTTCACAAAGAAAGGCATTTATGCAGCCCAAAACACATGAAAAA 29140
Db 810	AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLys***HisMetLysLys 829
QY 29139	TACTCATCATCTAGCCCATCAGAGAAATGCAAAATCAAAACCAATGAGATACCATCTC 29080
Db 830	CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 849
QY 29079	ACACGAGTTAGATGCATCATTTAAAAAGTCAGGAACAAACAGGTGCTGGAGAGATGT 29020
Db 850	ThrProValArgMetAlaIleIleLysSerGlyAsnAsnArgCysTrpArgGlyCys 869
QY 29019	GGAGAATAGGAACACTTTTACACTGTTGGTGGAGCTGTGAAGTCTCAACCATTTGG 28960
Db 870	GlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLeuTrp 889
QY 28959	AAGTCAAGTGTGTGATTCCTCAGGGATCTAGAACTAGAATACATTTGACCCAGCCATC 28900
Db 890	LysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIle 909
QY 28899	CAATTAAGTGGTATATACCAAGGATTATAATCATGTGCTATTAAGACACAAACGACA 28840
Db 910	ProLeuLeuGlyIleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThr 929
QY 29839	TGTATCTTTATAGCACACACTATTCACAATAGCAAAAGACTTGGAAACCACTTAATGTCCA 28780
Db 930	ArgMetPheIleAlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysPro 949
QY 28779	ACAACCATAGACTGGGATTAAGAAAATGTGGCACATATACACCATGAATACTATGCAGCC 28720
Db 950	ThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAla 969
QY 28719	ATAAAAAATGATGAGTTCGTGCTCTTTGTAGGACATGGATGAAGCTGGAACCATCATTT 28660
Db 970	IleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu***IleIle 989
QY 28659	CTCAGCAAACTATCAACAAGCACAAAACCAACACCGCATGCTCTCACTCATAGTGGG 28600
Db 990	LeuSerLysLeuSerGln***GlnLysThrLysHisArg***PheSerLeuIleGlyGly 1009
QY 28599	AAT 28597
Db 1010	Asn 1010
RESULT 8	
US-10-000-256A-153	
; Sequence 153, Application US/10000256A	
; Publication No. US20030039983A1	
; GENERAL INFORMATION:	
; APPLICANT: Sun, Yongming	
; APPLICANT: Recipon, Hervé	
; APPLICANT: Chen, Sei-Yu	
; APPLICANT: Liu, Chenghua	
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a	
; FILE REFERENCE: DEX-0259	
; CURRENT APPLICATION NUMBER: US/10/000, 256A	
; CURRENT FILING DATE: 2001-11-01	
; PRIOR APPLICATION NUMBER: 60/244,782	
; PRIOR FILING DATE: 2000-11-01	
; NUMBER OF SEQ ID NOS: 240	
; SOFTWARE: PatentIn version 3.1	
; SEQ ID NO 153	
; LENGTH: 956	
; TYPE: PRT	
; ORGANISM: Homo sapien	
US-10-000-256A-153	
Alignment Scores:	
Pred. No.:	1,45e-168
Score:	2187.00
Percent Similarity:	94.51%
Best Local Similarity:	91.76%
Query Match:	4.17%
Indels:	0
Gaps:	0
US-10-083-853B-2 (1-29921) x US-10-000-256A-153 (1-956)	
QY 29907	GAATTCGAAAAAAGTACTTTAAAGTTTCATGTGGAAACCAAAAAAGAGCCACATCGCCAAG 29848
Db 520	GluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAlaHisIleAlaLys 539
QY 29847	TCAATCCTAAGCCAAAAGAACAAAGCTGGAGGCATCATGCTACTGCTCAAACTACAC 29788
Db 540	SerIleLeuAsnGlnLysAsnLysAlaGlyIleThrLeuProAspPheLysLeuTyr 559
QY 29787	TACAAAGCTACAGTACCAACCAACAAACATGGTACTGGTACCAACAGAGATATCAACCA 29728
Db 560	TyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAspGln 579
QY 29727	TGGAACAGAACAGAGCCCTCAGAAATTAATGCCCATATCTACAACATATCTGATCTTTGAC 29668
Db 580	TrpAsnArgThrGluProSerGluIleThrGlnHisIleTyrSerTyrLeuIlePheAsp 599
QY 29667	AAACCTGAGAAAAACCAAGCAATGGGAAAGGATTCCTATTTAATAAATGGTGTGAGAA 29608

```
Db 600 LysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGlu 619
QY 29607 AACTGCCTAGCCTATGTAGAAAGCTGAAAGCTGAACTCCCTTACACCTTATACAAA 29548
Db 620 AsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLys 639
QY 29547 ATTAATCAAGATGATTAAAGACTTAATGTAGACCTAAACCAATAAAATCCCTAGAA 29488
Db 640 MetAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGlu 659
QY 29487 GAAACCCAGGCAATACCATTCAGGACATAGGATGGCAGGACTTCATGCTCTAAACA 29428
Db 660 GluAsnLeuGlyIleThrIleGlnAspIleGlyMetGlyLysAspPheMetSerLysThr 679
QY 29427 CCAAAAGCAATGGCAACAAAGCCAAATTTGCAAAATGGGATCTAAATTAATAAGAGC 29368
Db 680 ProLysAlaMetAlaThrLysAspLysIleAspLysTrpAspLeuValLysLeuLysSer 699
QY 29367 TTCTGCACGCAAAAGAACTACCATCAGAGTGAACAGGCAACTTACAGAAATGGGAGAA 29308
Db 700 PheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThrLysTrpGluLys 719
QY 29307 ATTTTGTCAACTACTCATTTGACAAAGGCTAATATCCAGAACTACAATGAACCTCAA 29248
Db 720 IlePheAlaThrTyrSerAspLysGlyLeuIleSerArgIleTyrAsnGluLeuLys 739
QY 29247 CAAATTTCAAGAAAGAAACCAAGAACCCCATCAAAAGTGGGTGAAGGATATGAACAGA 29188
Db 740 GlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAsnArg 759
QY 29187 CACTTCTCAAAGAGGCAATTTATGCAGCCCAAAACACATGAAAAATATCTCATCATCA 29128
Db 760 HisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSerSer 779
QY 29127 CTGGCCATCAGAGAATCCAAATCAAAACCAATGAGATACCATCTCACACCACTTAGA 29068
Db 780 LeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProValArg 799
QY 29067 ATGCAATCATTAAGTACAGGAACCAACAGGCTGCTGGAGAGATGGGAAATAGGA 29008
Db 800 MetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluThrGly 819
QY 29007 ACATTTTACACTGTGTGGGACTGTGAAGTGTCAACCATGTGTGAAGTCACTGTGG 28948
Db 820 ThrLeuLeuHisCysTrpTrpAspCysLysLeuAlaGlnProLeuTrpLysSerValTrp 839
QY 28947 TGATTCCTCAGGATCTAGAACTAGAAATACCATTTGACCCAGCATCCCATTTCTGGGT 28888
Db 840 ArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIleProLeuLeuGly 859
QY 28887 ATATCCCAAGGATTATAATCATGCTGCTATAAAGACACAAAGCACATGTATGTTTATA 28828
Db 860 IleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThrArgMetPheIle 879
QY 28827 GCAGCACTATTCAACATAGCAAGACTTGGAAACCACTAATGTCCAAACCAAGATAGAC 28768
Db 880 AlaAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProThrIleIleAsp 899
QY 28767 TGGATTAAAGAAATGTGGACATATACACCATGGAATACTATGAGCCATATAAAATGAT 28708
Db 900 TrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLysAsnAsp 919
QY 28707 GAGTTCTGCTCTTTGTAGGACATGAGTGAAGCTGGAAACCATCATCTTCAGCAAACTA 28648
Db 920 GluPheValSerPheValGlyThrTrpMetLysLeuGluIleIleLeuSerLysLeu 939
QY 28647 TCACAGGACAAAAACCAACCGCATGTTCTCACTCATAGTGGGAAT 28597
Db 940 SerGlnGluGlnLysThrHisArgIlePheSerLeuIleGlyGlyAsn 956
```

RESULT 9

US-10-001-835-189

; Sequence 189, Application US/10001835

```
; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 189
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-189
```

## Alignment Scores:

Pred. No.:	5.58e-110	Length:	917
Score:	1467.00	Matches:	275
Percent Similarity:	95.58%	Conservative:	6
Best Local Similarity:	93.54%	Mismatches:	13
Query Match:	2.80%	Indels:	0
DB:	9	Gaps:	0

US-10-083-853b-2 (1-29921) x US-10-001-835-189 (1-917)

```
QY 29919 ACTTCTTCCACAGAAATGGAAAAAATACTTTAAAGTTTCATGTGGAACCAAAAAAGAGCC 29860
Db 624 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 643
QY 29859 CACATCGCCAAAGTCAATCTTAAGCCAAAGAACAAAGCTGGAGGATCATCTACCTGAC 29800
Db 644 ArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProasp 663
QY 29799 TTCAAACTACACTCAAGGCTACAGTACCAAAACACATGCTACTGTGTACCAAAACAGA 29740
Db 664 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpGlnAsnArg 683
QY 29739 GATATCAACCAATGGAACAGACAGAGCCCTCAGAAAATAATGCCGCATATCTACAACTAT 29680
Db 684 AspIleAspGlnTrpAsnArgThrGluProSerGluIleMetProHisIleTyrAsnTyr 703
QY 29679 CTGATCTTTGACAAACCTGAGAAAACAAAGCAAGCAATGGGAAAGGATTCCTATTTAATAA 29620
Db 704 LeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLys 723
QY 29619 TGGTGTGAGAAAACTGGCTAGCCCTATGTAGAAGCTGAAACTGGATCCCTCTCTTACA 29560
Db 724 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThr 743
QY 29559 CCTTATACAAAATAATTAATTCAGATGGATTAAAGACTTAAATGTAGACCTTAAACCCATA 29500
Db 744 ProTyrThrLysIleAsnSerArgTyrIleLysAspLeuAsnValArgProLysThrIle 763
QY 29499 AAATCCCTAGAGAAAACCCAGGCAATACCATTCAGGCATAGGATGGGCAAGGACTTC 29440
Db 764 LysThrLeuGluLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspPhe 783
QY 29439 ATGTCTAAACCAACCAAGCAATGGCAACAAAGCCAAAATGACAAATGGGATCTAAT 29380
Db 784 MetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIle 803
QY 29379 AAATAAGAGCTTCTGCACAGCAAAAGAAACTACCATCAGAGTGAACAGGCAACTTACA 29320
Db 804 LysLeuLysSerPheCysThrAlaLysGluThrThrIleArgValAsnArgGlnProThr 823
QY 29319 GAATGGGAGAAAATTTTGGCAACCTACTCATTTGCAAAAGGGCTAATATCCAGAACTAC 29260
```

Db 824 ThrTrpGluLysIlePheAlaThrTyrSerAspLysGlyLeuIleSerArgIleTyr 843  
QY 29259 AATGAACCTCAACAAATTTACAGAAACAAACAAAGACCCCATCAAAAAGTGGGTCAAG 29200  
Db 844 AsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTyrPalaLys 863  
QY 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTATGACGCCCAAAAACACATGAAAAAA 29140  
Db 864 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLysLys 883  
QY 29139 TACTCATCATCTGGCCATCAGAGAAATCAAAATCAAAACCAACATGAGATCCATCTC 29080  
Db 884 CysSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeu 903  
QY 29079 ACACCACTTGAATGGCAATCAATAAAAGTCAGGAACAAC 29038  
Db 904 ThrProValArgMetAlaIleIleLysLysSerGlyAsnAsn 917  
RESULT 10  
US-10-082-830-235  
; Sequence 235, Application US/10082830  
; Publication No. US20030077604A1  
; GENERAL INFORMATION:  
; APPLICANT: Sun, Yongming  
; APPLICANT: Recipon, Hervé  
; APPLICANT: Salceda, Susana  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Turner, Leah  
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific  
; FILE OF INVENTION: Genes and Proteins  
; FILE REFERENCE: DEX-0249  
; CURRENT APPLICATION NUMBER: US/10/082,830  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,802  
; PRIOR FILING DATE: 2000-10-27  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 235  
; LENGTH: 940  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-082-830-235  
Alignment Scores:  
Pred. No.: 1.56e-100 Length: 940  
Score: 1351.00 Matches: 251  
Percent Similarity: 90.51% Conservative: 16  
Best Local Similarity: 85.08% Mismatches: 28  
Query Match: 2.58% Indels: 0  
Gaps: 0  
US-10-083-853B-2 (1-29921) x US-10-082-830-235 (1-940)  
QY 29919 ACTTCTTCACAGAAATTTGAAAAAAGTCTTAAAGTTCAATGTGGAACCAAAAAGAGCC 29860  
Db 646 ThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGlnLysArgAla 665  
QY 29859 CACATCCCAAGTCATCTTAAGCAAAAGAAACAAAGCTGGAGGCATCATCTACCTGAC 29800  
Db 666 ArgIleAlaLysThrIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAsp 685  
QY 29799 TTCAAACTACACTACAGAGGTACAGTACCAACCAACATGGTACTGGTACCAAAAACAGA 29740  
Db 686 PheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTyrTrpTyrGlnAsnArg 705  
QY 29739 GATATCAACCAATGGACAGACAGACGCCCTCAAAATAATGCGCATATCTACAACAT 29680  
Db 706 AspileAspGlnTrpAsnArgIleGluProLeuGluLeuIleProHisIleTyrAsnHis 725  
QY 29679 CTGATCTTTGACAAACCTGACAAAACAAACAAACATGGGGAAGATTCCTCTATTATAAA 29620  
Db 726 LeuIlePheAspLysProAspLysAsnLysLysLeuTrpGlyLysAspSerLeuPheAsnLys 745

QY 29619 TGTGCTGAGAAAAGCTAGCCCTATGTAGAAAGCTGAAAGTGGATCCCTTCCTTACA 29560  
Db 746 TrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnLeuPheLeuThr 765  
QY 29559 CCTTATACAAAATAATTAATCAAGATGGATTAAAGACTTAATGTGTAGACCTAAACCAT 29500  
Db 766 ProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIle 785  
QY 29499 AATCCCTAGAAAACCCAGCAATACCATTCAGACATAGGCATGGCAGGACTTC 29440  
Db 786 LysIleLeuGluLysAsnLeuGlyAsnThrIleGlnAspIleGlyValGlyLysAspPhe 805  
QY 29439 ATGCTTAAACACCAAAAGCAATGGCAACAAAGCCAAAATTCACAAATGGGATCTAAT 29380  
Db 806 MetThrLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTyrPaspIleIle 825  
QY 29379 AACTAAAGAGCTTCTGCACAGCAAAAGAACTACCATCAGAGTGAACAGCAACTTACA 29320  
Db 826 LysLeuLysSerPheCysThrAlaLysGluThrThrIleIleValAsnArgGlnProThr 845  
QY 29319 GAATGGAGAAAATTTTTCACACTTACTCATTTGACAAAAGGCTTAATATCCAGAACTAC 29260  
Db 846 GluTrpGluLysIlePheLysIleTyrProSerAspLysGlyLeuIleSerArgIleTyr 865  
QY 29259 AATGAACCTCAACAAATTTTACAAAGAAAACCAAGAACCCCATCAAAAAGTGGTGAAG 29200  
Db 866 LysGluLeuLysGlnIleTyrLysLysLysSerAsnAsnProIleLysAsnTrpAlaLys 885  
QY 29199 GATATGAACAGACACTTCTCAAAAGAGGCAATTTATGACGCCCAAAAACACATGAAAA 29140  
Db 886 AspMetAsnArgHisPheSerLysGluAspIleTyrAlaValAsnArgHisMetLysThr 905  
QY 29139 TACTCATCATCTGGCCATCAGAGAAATCAAAAGTCAATCAAAACCAACATGAGATCCATCTC 29080  
Db 906 CysSerSerLeuLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisPhe 925  
QY 29079 ACACCACTTGAATGGCAATCAATAAAAGTCAGGAACAACAGG 29035  
Db 926 ThrProValArgMetAlaSerIleLysLysSerGlyAsnAsnArg 940

RESULT 11  
US-10-025-201-2  
; Sequence 2, Application US/10025201  
; Publication No. US20030003468A1  
; GENERAL INFORMATION:  
; APPLICANT: Crow, Mary K.  
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY  
; FILE REFERENCE: 5983/2H567  
; CURRENT APPLICATION NUMBER: US/10/025,201  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,673  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 338  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank  
; DATABASE ENTRY DATE: 1995-02-02  
; RELEVANT RESIDUES: (1)..(338)  
US-10-025-201-2  
Alignment Scores:  
Pred. No.: 1.48e-79 Length: 338  
Score: 1090.50 Matches: 234  
Percent Similarity: 79.18% Conservative: 36  
Best Local Similarity: 68.62% Mismatches: 67  
Query Match: 2.08% Indels: 6  
Gaps: 1  
DB:

US-10-083-853B-2 (1-29921) x US-10-025-201-2 (1-338)

[illegible]

RESULT 12									
US-10-083-853-2									
; Sequence 2, Application US/10083853									
; Patent No. US20020164709A1									
; GENERAL INFORMATION:									
; APPLICANT: Affymetrix, Inc									
; APPLICANT: Shigetata, Ron T									
; APPLICANT: Siani-Rose, Michael A									
; TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein									
; FILE REFERENCE: 3385.1									
; CURRENT APPLICATION NUMBER: US/10/083,853									
; CURRENT FILING DATE: 2002-02-26									
; PRIOR APPLICATION NUMBER: USSN 60/272,663									
; PRIOR FILING DATE: 2001-03-01									
; NUMBER OF SEQ ID NOS: 2									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 2									
; LENGTH: 317									
; TYPE: PRT									
; ORGANISM: Homo Sapiens									
US-10-083-853-2									
Alignment Scores:									
Pred. No.:		2.04e-56		Length:		317			
Score:		806.00		Matches:		182			
Percent Similarity:		51.88%		Conservative:		11			
Best Local Similarity:		48.92%		Mismatches:		25			
Query Match:		1.54%		Indels:		154			
DB:		9		Gaps:		6			
US-10-083-853B-2 (1-29921) x US-10-083-853-2 (1-317)									
Qy	29454	ATGGC	CAAGG	ACTCAT	GTCTCT	CTAAAC	ACACAA	AGCAAT	GGCAACAAAGCCAA
Dd	1	MetGly	LysAsp	PhMet	SerLys	ThrPro	LysAla	MetAla	ThrLysAla
Qy	29394	AAATG	GGATCT	CAATTA	AACTTA	AGAGCT	TCTG	CACACG	CAAAAGAAACTACCA
Dd	21	LysTrp	AspLeu	LeuLys	SerPhe	CysThr	AlaLys	GluThr	ThrTr
Qy	29334	AACAG	GCAACT	TACAGA	TGGAGA	AAATTT	TTCG	CAACCT	ACTCATTTGACA
Dd	41	AsnArg	GlnLeu	ThrGlu	TrpGlu	LysIle	PheAla	ThrTyr	SerPheAsp
Qy	29274	ATATC	CAGAAT	CTACAAT	GAACTCA	AAACAA	ATTTTACA	AGAAAA	AAAAACAAAGAA
Dd	61	IleSer	ArgIle	TyrAsn	GluLeu	LysGln	IleTyr	LysLys	LysThrLysAla
Qy	29214	AAAAA	TGGTGA	AGGATAT	GAACAG	ACACTT	CTCTCA	AAAGAG	CGCATTTATG
Dd	81	LysLys	TrpVal	LysAsp	MetAsn	ArgHis	PheSer	LysGlu	GlyIleTyrA
Qy	29154	AAACA	CATG	AAAAAT	ACTCAT	CATC	ATCG	CGCCAT	CAGAGAAATGCAAAATCA
Dd	101	LysHis	MetLys	LysTyr	SerSer	SerLeu	AlaIle	ArgGlu	MetGlnIleL
Qy	29094	ATGAG	ATACCA	CTCTCAC	ACCAGT	TAGAAT	TGGCAAT	CATTAAAA	AGTCAGGAA
Dd	121	MetArg	TyrHis	LeuThr	ProVal	ArgMet	AlaIle	IleLys	LysSerGlyA
Qy	29034	TGCTG	GAGAG	ATGTGG	AGAAAT	AGGAAC	ACITTTT	TACACT	TGTTGGTGGGACT
Dd	139	-----							
Qy	28974	GTTCA	ACCATT	TGTGGA	AGTCA	GTTGGT	GATTCT	CCTCAG	GATCTAGAACTAG
Dd	139	-----							
Qy	28914	TTTGAC	CCAGCA	ATCCCA	TATTCT	GGGTAT	ATAC	CCCAAG	GATTATTAATCAT
Dd	139	-----							

```
QY 28854 AAAGACACAGCACATGTATGTTTATAGCAGCACTATTACAAATAGCAAAAGACTTGGAAAC 28795
Db 139 ----- 139
QY 28794 CAACCTAAATGTCCTCAACAACAGTAGACTGGATTGAAGAAATGTGGCACATATACACCATG 28735
Db 139 ----- 139
QY 28734 GAATACTATGCAGCCATAAAAAATGATGAGTTCGTGCTTTGTAGGGACATGGATGAAG 28675
Db 140 -----ArgAspMeCAspGluA 145
QY 28674 CTGGAACCATCATCTCTCAGCAAACTATCACAAGGACAAACCAACCCGATGTC 28615
Db 145 laglyAsnHisSerGlnGlnThrIleThrArgThrLysAsnGlnThrProHisVal 165
QY 28614 TCACTCATAGGTGGGAATTGAACAATGAGAACACATGGACACAGGAGGGGAACATCACA 28555
Db 165 eutrHisArgTrpIleLeu---GlnGlnSerHisTrpVal----- 177
QY 28554 CACTGAGGACTGTGTGGGTGGGGAGGGGGAGGATAGCATTTAGGAGATATACCTA 28495
Db 178 -----ThrValLeuSerAspIleSerG 185
QY 28494 ATGTTAAATGACGAGTTAATGGGTGCACACACCAACATGGCACATGTATACATATGTA 28435
Db 185 luLeu-----MetHisLysThrAspArgIleV 194
QY 28434 CAACCTGCACATGTGCATATGTACTCTAAACTTAAAGTATAATAATCAATAAATAA 28375
Db 194 alAsnLeu-----LeuMetCysMetTyr-----LeuLeuThrValAspLeuA 208
QY 28374 AAAAAACAAACAAACAGAAACACATATACC 28343
Db 208 spArgLeuAsnAspAspAlaLysArgTyrSer 218
```

## RESULT 13

US-10-001-876-169  
; Sequence 169, Application US/10001876  
; Patent No. US2002017140A1

## GENERAL INFORMATION:

; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Hervé  
; APPLICANT: Cafferkey, Robert  
; APPLICANT: Ali, Shujath  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and  
; FILE REFERENCE: DEX-0285  
; CURRENT APPLICATION NUMBER: US/10/001,876  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,186  
; PRIOR FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 211  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 169  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Homo sapien  
; US-10-001-876-169

## Alignment Scores:

Pred. No.:	2,88e-53	Length:	341
Score:	767.50	Matches:	190
Percent Similarity:	48.37%	Conservative:	48
Best Local Similarity:	38.62%	Mismatches:	88
Query Match:	1.47%	Indels:	167
DB:	9	Gaps:	8

US-10-083-853B-2 (1-29921) x US-10-001-876-169 (1-341)

QY 4565 AATAGACCACTAGACAGACTAATAAGAGAGAGAGAGAGAAATCAAGAGATGCAATA 4624

```
Db 11 AsnLysProLeuAlaGluLeuArgLysLysArgLys-----Ile 25
QY 4625 AAAAAATGATAAGGGGATATCACCACCGATCCACAGAAATACAAACTATTATCAGAGAA 4684
Db 26 ArgAsnGluArgGluAspIleThrGluProThr---IleLysLysAsnIleAsnGlu 44
QY 4685 TATTATAAACACCTCTATGCAATAAATAGAAAATCTAGAAAATGGAATAATTCCTG 4744
Db 45 TyrTyrGluAlaLeuHisIleAsnGluLeuAspAsnLeuGluMetGluLysPheLeu 64
QY 4745 GACACATATGTAGCCTGTATGGACCTTGGGGGAGAGACAAAAGGGGTGAATGCAGAA 4804
Db 65 ThrIleTyr----- 67
QY 4805 TAAACACAAAGACAAAGAGATGTGTTGGAAGTAGGGGTCAGGGGGCAACTTGCCTCTA 4864
Db 67 ----- 67
QY 4865 ATGGACAAGGGCCCTGAGCTTTACACACCCCTCTGTATTATTAGCAAAAGAGATAGCG 4924
Db 67 ----- 67
QY 4925 AGAGGCTGAGTTGGAAGAAGAGTCAGCTGTTAGGTCCAGAGTAGCCCTGCAAGACTGCA 4984
Db 67 ----- 67
QY 4985 TTCCTCAACAATAGGCTCTAGATGTCACAGTAGATAACCTCAAGGAGCCAGTGCAGGG 5044
Db 68 -----AspLeuPro----- 70
QY 5045 AGTGATGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAAGTAAGTTTGGCCCACTCTG 5104
Db 70 ----- 70
QY 5105 TATTCACGATAAACAGTTTGTCTGTTGATCAAGTAGCCTCCAGTGAATGCTGAGTTGGT 5164
Db 70 ----- 70
QY 5165 CATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCCTCTCAAGACTAAACCAG 5224
Db 71 -----LysGln 72
QY 5225 GAAGAAGTCAAAATCCCTGAATATACAGTAACAGTCTTAAAATTTGAAGCAGTAATTGAT 5284
Db 73 GluValThrGluAsnLeuAsnLysProIleThrSerHisGluThrAlaValArgIleLys 92
QY 5285 AGCCTTACCACCAACCAAAAAAGTCAGGACGAGGATTCACAGCCAAATTTCTACCAGAGG 5344
Db 93 LysLeuProValLysLysSerProGlyGlnAspGlyPheIleSerLeuPheAlaGlnThr 112
QY 5345 TACAAGAGAAAGCTGTGCTACTATTCTCTGAAACTATTCCAAAAATA---GAAATGGG 5401
Db 113 PheLysGluGluLeuIleProIleLeuLeuLysLysPheGlnLysIleGluGluGlu-Gl 132
QY 5402 AATCCTCCCTAACTCATTTTACGAGGCCAGCATCATCTGATACCAAAACCTAGCAGTGA 5461
Db 132 ylleLeuProAsnSerPheTyrLysAlaSerIleThrLeuIleProLysProAspLysAs 152
QY 5462 CACAACA-----AAAAGAGGAAATTTTCAGGCCCATATCCCTGATGATCAACATTCATGT 5512
Db 152 pThrSerLysIleIleLysLysAlaAsnTyrArgProIleSerLeuMetAsnThrAspAl 172
QY 5513 GAAATCTCATAATAAATCTGCAAAACCAAAATCCAGCAGCAGCATCAAAAAGCTTATCTA 5572
Db 172 alYsIleLeuAsnLysMetLeuAlaAsnHisIleGlnGlnTyrIleLysLysIleIleH 192
QY 5573 CCATGATCAAGTTGGCGTTCATCCCTGGGATGCAAGGCTGGTTCAAAATATGCAAAATCAAT 5632
Db 192 SHISAspGlnValGlyTyrValProGlyMetGlnGlyTrpPheAsnIleCysLysSerIl 212
QY 5633 AAATGTAGGCCCACTACATAAACAGAACCAATGACAAAAACCAACATGATTATCTCAATAGA 5692
```



Publication No. US20030092898A1  
GENERAL INFORMATION:  
APPLICANT: Salceda, Susana  
APPLICANT: Macina, Roberto  
APPLICANT: Hu, Ping  
APPLICANT: Recipon, Herve  
APPLICANT: Karra, Kalpana  
APPLICANT: Cafferkey, Robert  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific  
FILE OF INVENTION: Genes and Proteins  
FILE REFERENCE: DEX-0313  
CURRENT APPLICATION NUMBER: US/10/074,475  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 60/268,292  
PRIOR FILING DATE: 2001-02-13  
NUMBER OF SEQ ID NOS: 295  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 250  
LENGTH: 190  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-074-475-250

Alignment Scores:  
Pred. No.: 6e-46 Length: 190  
Score: 676.00 Matches: 147  
Percent Similarity: 82.65% Conservative: 15  
Best Local Similarity: 75.00% Mismatches: 25  
Query Match: 1.29% Indels: 10  
DB: Gaps: 3

US-10-083-853B-2 (1-29921) x US-10-074-475-250 (1-190)

QY	4151	TTAGAACTCAGAATTAGAACTCCTCAAAATCACAACTACATGGAACTGAACAAC	4210
Db	3	LeuGlnLeuArgIleLysSerLeuThrGlnAsnArgThrThrTrpLysLeuAsnAsn	22
QY	4211	CTGCTCTGTAATCAGTACTGGTAAATAACAAATGAAGGCAAAATAAAGATGTCCTTT	4270
Db	23	LeuLeuLeuAsnAspTyrTrpValAsnLysLysIleLysAlaGluIleAsnLysPhePhe	42
QY	4271	GAACCAATGAGAACAAAGACAAATGTACAGAACTCTCGGGCATATTTAAAGCAGTG	4330
Db	43	GluThrIleGluAsnLysAspThrMetTyrGlnAsn-----ThrAlaLysAlaVal	59
QY	4331	TGTAGAGGGAATTTATAGACTAGATGCTTACAGAGAAAGCAGGAATATCTAAATA	4390
Db	60	PheArgGlyLysPheIleAlaLeuAsnThrHisIleArgAsnTrpGluIleProLysIle	79
QY	4391	GACACCTTAACATCACAATTTAAAGAACTAGAGAGAAAGACAAACAATTCAAAAGCT	4450
Db	80	AsnValLeuThrSerGlnLeuLysGluLeuLysArgGluGlnThrHisSerLys---	98
QY	4451	AGCAGAAGACAGAAATAACTAAGATCAGAGCAAGCTGAAGGAGATAGACACAAAAA	4510
Db	99	-----GlnGluIleThrLysIleIleAlaGluLeuLysGluIleGluThrGlnLys	115
QY	4511	GCCTTCAAATAATCAATGAATCCAGGAGCTGGTTTTTTGAAAAGATCAGCAAAATAG	4570
Db	116	AlaLeuGlnLys-IleSerAspSerArgSerTrpPheGluLysIleAsnLysThrAs	135
QY	4571	CCA---CTAGACAGATAAAGAAAGAAAGAGAGAAATCAAGAGATCAATAAA	4626
Db	135	PArgLeuLeuAlaArgIleIleLysLysLysArgGluLysAsnGlnIleAspThrIleLys	155
QY	4627	AAATGATAAGGGGATATCACCACCGATCCACAGAAATACAAATATTATCAGAGAATA	4686
Db	155	sasnAspLysGlyAspIleThrThrAsnProThrGluIleGlnThrAlaIleArgGluCy	175
QY	4687	TTATAAACCTCTATGCAAAATAACTAGAAAATCTAGAAGAAATG	4732

Db 175 sTyrGlnHisLeuTyrIleAsnLysLeuGluAsnLeuGluGluIle 190  
Search completed: July 4, 2003, 23:32:11  
Job time : 761.5 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 4, 2003, 17:35:48 ; Search time 625.5 Seconds  
(without alignments)  
12748.175 Million cell updates/sec

Title: US-10-083-853b-2

Genetic score: 52528  
Sequence: 1 gtaattgtaagaagcctca.....caattctgtgaagaagtaa 29921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgr2\_1/USFTO.spool/US10083853/runat\_03072003\_093611\_8069/app\_query.fasta\_1.30087  
-DB=A\_Geneseq\_101002 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US10083853 -ECGN\_1\_1\_1002 -runat\_03072003\_093611\_8069 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPHLOC=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	5266.5	10.1	2747	22	ABG02342 Novel human diagno
2	5244	10.0	1280	22	AAB64943 Human secreted pro
3	5164	9.9	1284	22	ABG06053 Novel human diagno
4	5164	9.9	1284	22	ABG09636 Novel human diagno
5	5164	9.9	1284	22	ABG10795 Novel human diagno
6	5164	9.9	1284	22	ABG14989 Novel human diagno
7	5086	9.7	1726	22	ABG14545 Novel human diagno
8	5067.5	9.7	2030	22	ABG24249 Novel human diagno
9	5000	9.6	2764	22	ABG10129 Novel human diagno
10	4989.5	9.5	1824	22	ABG09152 Novel human diagno
11	4902	9.4	2921	22	ABG08447 Novel human diagno
12	4893.5	9.4	1691	22	ABG03983 Novel human diagno
13	4893.5	9.4	1691	22	ABG06058 Novel human diagno
14	4893.5	9.4	1691	22	ABG08641 Novel human diagno
15	4893.5	9.4	1691	22	ABG10802 Novel human diagno
16	4869	9.3	4318	22	ABG14787 Novel human diagno
17	4858	9.3	1573	22	ABG23673 Novel human diagno
18	4811.5	9.2	1328	22	ABG10827 Novel human diagno
19	4800	9.2	1319	22	AAO12666 Human polypeptide
20	4797.5	9.2	1328	22	ABG09567 Novel human diagno
21	4785	9.1	3444	22	ABG17495 Novel human diagno
22	4784	9.1	1658	22	ABG06940 Novel human diagno
23	4782.5	9.1	1292	22	ABG08446 Novel human diagno
24	4698	9.0	1261	22	ABG07064 Novel human diagno
25	4642	8.9	2009	22	ABG06085 Novel human diagno
26	4638	8.9	1680	22	ABG17645 Novel human diagno
27	4619	8.8	1702	22	ABG06629 Novel human diagno
28	4582.5	8.8	1254	22	ABG19921 Novel human diagno
29	4573.5	8.7	1289	22	ABG18498 Novel human diagno
30	4537.5	8.7	3089	22	ABG12751 Novel human diagno
31	4527.5	8.7	2664	22	ABG02343 Novel human diagno
32	4508.5	8.6	1325	22	ABG14552 Novel human diagno
33	4501.5	8.6	1257	22	ABG17335 Novel human diagno
34	4489.5	8.6	1534	22	ABG14953 Novel human diagno
35	4483.5	8.6	2775	22	ABG14956 Novel human diagno
36	4425	8.5	1457	22	ABG06066 Novel human diagno
37	4425	8.5	1457	22	ABG09647 Novel human diagno
38	4425	8.5	1457	22	ABG17624 Novel human diagno
39	4425.5	8.5	1843	22	ABG14961 Novel human diagno
40	4418	8.4	1241	22	ABG14773 Novel human diagno
41	4418	8.4	1284	22	ABG14934 Novel human diagno
42	4395	8.4	1709	22	ABG02341 Novel human diagno
43	4395	8.4	1709	22	ABG03052 Novel human diagno
44	4395	8.4	1709	22	ABG03979 Novel human diagno
45	4395	8.4	1709	22	ABG06055 Novel human diagno

ALIGNMENTS

RESULT 1  
ABG02342  
ID ABG02342 standard; Protein; 2747 AA.  
XX ABG02342;  
AC ABG02342;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #2333.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX



Db	1528	oasnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAs	1548	Db	1845	-----	1845
QY	3841	CTCCACACAATAAGTGGAGTCTAAATAATAATAGACACTTTAAACACCCACTGCC	3900	QY	4917	AGATACGAGAGGGTCACTTGGAAAGAGGTCAGCTGTTAGTCCAGAGTAGGCTGCA	4976
Db	1548	pserHisThrLeuIleMetGly-----AspPheAsnThrProLeuSe	1562	Db	1845	-----	1845
QY	3901	AATATAGGACGATCAATCAGACAGAAAATTAACAGGATATCCAGGAGTGAACCTGAGC	3960	QY	4977	AGACTGCATTCCTCAACAATAAGGCTCTAGATGTCCAGTAGATAACCTCAAGGAGCCAG	5036
Db	1562	rThrLeuAspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeuAsnSerAl	1582	Db	1845	-----	1845
QY	3961	TCGGACCAAGCGGACCTAATATAGATATCTACAGAACTCCCCACCCAAATCAACAGAATA	4020	QY	5037	TGCCAGGGAGTGTATGCCCTCAGCAAACTTCTAGGGCAGGCACAGAACTAAGTTGGCC	5096
Db	1582	aLeuHisGlnAlaAspLeuIleAspIleThrArgThrLeuHisProLysSerThrGlutY	1602	Db	1845	-----	1845
QY	4021	TACACTCTCTCAGATCAGATACACCTATTTAAATTTGACCATGTAATTTAAGTAA	4080	QY	5097	ACATTCTGTATTACAGATAAACAGTTTGCTGTTGTATCAAGTAGCCTCCAGTGAATGCT	5156
Db	1602	rThrPheSerAlaProHisHisSerThrLysIleAspHisIleLeuGlySerGI	1622	Db	1845	-----	1845
QY	4081	AACACTCTCAGCAAAATGCAAAAGAACAGAAATCTTAACAAACAGTCTCTCAGACTACAG	4140	QY	5157	GAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACACCCCTCTCAAGAC	5216
Db	1622	uAlaLeuLeuSerLysCysLysArgThrGluIleIleThrAsnTyrLeuSerAspHis	1642	Db	1846	-----	1846
QY	4141	TCGAATCTATTAGAACTCAGAAATTAAGAACTCTACTCAAAATCACAACTACATGGAA	4200	QY	5217	TAAACGAGGAAGTCAAAATCCCTGAATATATACCAAGTAACTTCTTAAATTTGAAGCAG	5276
Db	1642	rAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGlnSerArgSerThrTriply	1662	Db	1852	euAsnGlnGluValGluSerLeuAsnArgProIleThrGlyAlaGluIleValAlaI	1872
QY	4201	ACTGAACAACCTGCTCTGAATGACTACTGGCTTAATACAAATCAAGGCAAAATATAA	4260	QY	5277	TAATTGATAGCTTACCACCAACCAAAAGTCCAGGACGAGGATTCACAGCAAAATTCCT	5336
Db	1662	sLeuAsnLeuLeuLeuAsnAsprrtrpValHisnGluMetLysAlaGluIleL	1682	Db	1872	leilEasSerLeuSerThrLysLysSerProGlyProAspGlyPheThrAlaGluPhe	1892
QY	4261	GATGTTCTTTGAAACCAATGAGAACAGACACAATGTACCAAGATCTCTGGGGCATATT	4320	QY	5337	ACCAGAGGTACAAAGAGAGCTGGTACTTCTTCTTGAACATTTTCCAAAAATAGAA -	5395
Db	1682	sNetPheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTripAspAlaPh	1702	Db	1892	yrGlnArgTyrLysGluGluLeuValProPhePheLeuLysLeuPheGlnSerIleGluL	1912
QY	4321	TAAAGAGTGTAGAGGGAATTTATAGCACTAGATGCCTACAGAGAAACGAGAAAT	4380	QY	5396	--AATGGGATCTCTCCCTAACTCATTTTACGAGGCGCAGCATCATCTGATACCAAACT	5453
Db	1702	eLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluAr	1722	Db	1912	ysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleLeuIleProLysPro	1931
QY	4381	ATCTAAATPAGACACTTAACATCACAAATTAAGAACTAGAGAAAGAGCAACAA	4440	QY	5454	AGCAGTGCACACAAACAAAGAGCAATTTTCAGGCGCATATCCCTGATGAACATTTGATG	5513
Db	1722	gserLysIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluThrHi	1742	Db	1932	GlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAla	1951
QY	4441	TTCAAAAGCTAGCAGAAGCAAAATAACTAAGATCAGAGCAGAACTGAAGGAGATAGA	4500	QY	5514	AAAATCCTCAATATAATACTGGCAACCAATCCAGCAGCACATCAAAAGCTTTATCTAC	5573
Db	1742	sSerLysAlaSerArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGI	1762	Db	1952	LysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHis	1971
QY	4501	GACAAAAAGCCCTTCAATTAATCAATGAATCAGAGCTGGTTTTTGAAGAAGATCA	4560	QY	5574	CATGATCAACTGGCGGTATCCCTGGATGCAAGCTGCTTCAAAATATGCAAAATCAATA	5633
Db	1762	uThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePheGluArgIleA	1782	Db	1972	HisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIle	1991
QY	4561	GCAAAAT---AGACCACCTAGACAGACTAATAAAGAAAGAGAGAGAAATCAAGAG	4616	QY	5634	AATGTAGGCCATCACATAAACAGAACCAATGACAAAAACACATGATTTATCTCAATAGAT	5693
Db	1782	snLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleA	1802	Db	1992	AsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleValSerIleAsp	2011
QY	4617	ATGCAATAAAAAATGATAAGGGGATATCACACCCATCCACAGAAATACAAACTATTA	4676	QY	5694	GCAGAAAAGCCCTTGTCAAAATTCACAGCCCTTCATGCTTCAAAATTCCTAGTAACTA	5753
Db	1802	spThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrI	1822	Db	2012	AlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeu	2031
QY	4677	TCAGAGAATTTATTAACACCTCTATGCAAAATAACTAGAAAATCTAGAGAAATGGATA	4736	QY	5754	GGTATCGATGGATGTATCTCAAAATTAAGAGCTATTATACAAACCCACAGCAATA	5813
Db	1822	leArgGluTyrThrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAsp	1842	Db	2032	--Val-----	2032
QY	4737	AATTCCTGGACACATATGTAGCTGTATGGACCTTTGGGGGACAGAACAAAGGGGTGA	4796	QY	5814	TCATACTGAATGGCAAAAACCTGGAGCATTCCTTTTGAGAACTGGCAACAGCAAGGAT	5873
Db	1842	hrPheLeuAsp-----	1845	Db	2032	-----	2032
QY	4797	TGCAGAAATAAAAAGACAAAGAGAGATATGTTTGGAAAGTAGGGGTGAGGGGCAACT	4856	QY	5874	GCCCTCTCTACCACTCCTTATTCAGATACTATTGGAGTTCTTGCCAGGGCAATCAGCG	5933
Db	1845	-----	1845	Db	2033	-----	2033
QY	4857	TGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCACCTCTGTATTTANTTAGGCAAAAG	4916	QY	5934	AATAGAAAGAAATAAGGGTATTTCAAATAGAAAGAGAGGAGTCTATTTGCTCTGTGTTG	5993
Db				Db	2042	InclLysGluIleLysGlyIleGlnLeuGlyLysGluGluValLysLeuSerLeuPhe	2062



XX

The polynucleotide sequences given in AAF33213 to AAF33261 encode the human secreted proteins given in AAB64882 to AAB64930. AAB64931 to AAB64991 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunomodulatory; antisclerotic; dermatological; immunosuppressive; anti-inflammatory; anti-HIV; immunostimulant; cytostatic; cardiac; vascular; antimicrobial; anti-angiogenic; ophthalmological; neuroprotectant; anticonvulsant; nootropic; antialzheimers; antiparkinsonian; and vulnerary. The polynucleotides and polypeptides can be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. Disorders that may be prevented, diagnosed and/or treated by the above methods include immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immuno-deficiency virus (HIV), infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 1280 AA:

Alignment Scores:  
 Pred. No.: 0 Length: 1280  
 Score: 5244.00 Matches: 1068  
 Percent Similarity: 78.90% Conservative: 69  
 Best Local Similarity: 74.12% Mismatches: 138  
 Query Match: 10.02% Indels: 173  
 DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x AAB64943 (1-1280)

QY	3413	CCAGCTAGTGTGCATATGCGAGGATCAAAATTCACACATAATAATTAACCTTAATGTA	3472
Db	1	ProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrIleLeuThrLeuAsnVal	20
QY	3473	ANTGGCTAAATCCCAATTAAGACACACAGACTGGCAAAATGGATAAAGAGTCAAGAC	3532
Db	21	AsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnAsp	40
QY	3533	CCATCAGTGTGCTATTACGAGGCGCCATCTCACATGAAGAAGACACATAGGCTCAA	3592
Db	41	ProSerValCysCysIleGlnGluThrHisLeuThrCysArgAspThrHisArgLeuLys	60
QY	3593	ATAAAGGGATGGAGGAAGATTTACCAAGTAAATGGAACCAAAACAAAAAAGCAGGGTT	3652
Db	61	IleLysGlyTrpArgIleIleThrGlnAlaAsnGlyLysGln--LysLysAlaGlyVal	79
QY	3653	GCATCTAGTCTCTGATTAACAGACTTTAAACCAACAAAGATCAAAAGACAAAGAA	3712
Db	80	AlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGlu	99
QY	3713	GGCATTACATATGTTAAGGCATCAATGGACAGAGAGCTAACTATCTTAATATA	3772
Db	100	GlyHisTrpIleMetValLysGlySerIleGlnGlnGluLeuThrIleLeuAsnIle	119
QY	3773	CATGCACCAATACAGGACGCCAGATTCAATAAGCAAGTCTTAGAGACCTACAAAGA	3832
Db	120	TyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArg	139
QY	3833	GACTTTGACTCCACAAATAATAGTGGAGTCTAAATAATAATAGACACTTTACACC	3892
Db	140	AspLeuAspSerHisThr**IleMetGly-----AspPheAsnThr	153
QY	3893	CCACTGCCAATATTAGGCAGATCAATGACACAGAAAATTAACAGGATATCAGGAGTTG	3952
Db	154	ProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAspThrGlnGluLeu	173

QY	3953	AACTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCCACCCCAATCA	4012
Db	174	AsnSerAlaIleuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisProLysSer	193
QY	4013	ACAGAATATACACTCTCTCAGCATCACAATACACCTATTTTAAATTTGACCATGTAAT	4072
Db	194	ThrGluTyrThrPhePheSerAlaProHisHisThrTyrSerLysIleAspHisIle***	213
QY	4073	TAAAGTAAACACTCTCTCAGCAAAATGCCAAAGAACAGAAATCTCTACAACACTCTCTCA	4132
Db	214	GlySerLysAlaLeuLeuSerLysCysLysArgThrGluIleIleThrAsn**LeuSer	233
QY	4133	GACTCAGTGCATCTTATTAGAACTCAAGATTAAGAAATCTACTCAAAATCACAACT	4192
Db	234	AspHisSerAlaIleLysLeuGluLeuArgIleLys**LeuThrGlnAsnArgSerThr	253
QY	4193	ACATGGAATGAACAACTGCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4252
Db	254	ThrTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMetLysAla	273
QY	4253	AAATAAAGATGCTCTTGAACCAATGAGAAACAGACACATGTACCATGCTCTCTG	4312
Db	274	GluIleLysMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTyrp	293
QY	4313	GGCATATTTAAAGCAGTGTGTAGAGGGAATTTATAGCACTAGATGCCTACAGAGAAAG	4372
Db	294	Asp**PheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLys	313
QY	4373	CAGGAATATCTAAATAGACACCTTAAATCAATCAATTAAGAACTAGAGAAAGAG	4432
Db	314	GlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlnGluLys	333
QY	4433	CAACAAATCAAAAGCTAGCAGACAGAAATCAATCAATCAATCAATCAATCAATCAATCA	4492
Db	334	GlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLys	353
QY	4493	GAGATAGACACACAAAGCCCTTCAATTAATCAATCAATCAATCAATCAATCAATCAAT	4552
Db	354	GluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPheGlu	373
QY	4553	AAAGATCAGCAAAAT---AGACCCTAGACAGACTAATAAGAAAGAAAGAGAGAA	4608
Db	373	u***IleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgLysLys	393
QY	4609	TCAAAGAGATGCAATAAAAAATGATAAGGGGATATCACCCGATCCACAGAAATACA	4668
Db	393	nGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrAspProThrGluIleGlu	413
QY	4669	AACTATTATCAGAGAAATATTATAACACCTCTATGCAAAATAAACTAGAAAATCTAGA	4728
Db	413	nThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGlu	433
QY	4729	ANTGGATAAATTCCTGGACACATATGACCTGTATGGACCTTGGGGGAGAGAAAG	4788
Db	433	uMetAspLysPheLeuAsp-----	439
QY	4789	GGGTGAATGCAGAAATAAAGACAAAGAGATGTTTGGAAAGTAGGGGTGAG	4848
Db	439	-----	439
QY	4849	GGGCAACTTGCCTTAATGGACAAGGCCCTGAGCTTTACACCCTCTGTATTATTA	4908
Db	439	-----	439
QY	4909	GGCAAAAGAGATAGCGAGAGGTGAGTTGGAAGAGAGGTGAGTCTTGGTCCAGAGTA	4968
Db	439	-----	439
QY	4969	GGCTGCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCCAAGTAGATAACCTCA	5028
Db	439	-----	439

QY 5029 GGAGCCAGTGCAGGGAGTGTATGCCCTCAGCAAAACCTCTTAGGGCAGGCACAGAGTAA 5088  
Db 439 ----- 439  
QY 5089 GTTTGCCACATTCGTATTACGATAAACAGATTGTGTTGATCAAGTAGCCCTCCAGT 5148  
Db 439 ----- 439  
QY 5149 GGAATGCTGAGTTGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACATACACCT 5208  
Db 440 -----ThrTyrThrLe 443  
QY 5209 CTCAGAGCTAAACAGGAAGAGTCAATCCCTGAATATACCAAGTAAAGTCTCAAAAT 5268  
Db 443 uProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGlySerGluI 463  
QY 5269 TGAAGCAGTAATTCATAGCCCTACCAACCAAAAAAGTCCAGGACACAGCGATTACAGC 5328  
Db 463 e\*\*\*AlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrAl 483  
QY 5329 CAAATTTCTACAGAGGTACAAGAGAGCTGGTACTATTCTCTGAAACTATTCCAAAA 5388  
Db 483 aGluPheTyrGlnArgTyrLysGluLeuValProPheLeuLeuLysLeuPheGlnSe 503  
QY 5389 AATAGAA---AATGGGAATCCCTCACTCACTATTTAGAGGCCAGCATCTCTGATAC 5445  
Db 503 rIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleP 523  
QY 5446 CAAACCTTAGCAGTGACACACAAAAAGAGGAATTTCCAGCCCATCTCCCTGATGAACA 5505  
Db 523 rOlysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnI 543  
QY 5506 TTGATGTGAAATCCTCAATAAAATACTGGCAAAACCAATCCAGCAGCACATCAAAAAAGC 5565  
Db 543 leAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysL 563  
QY 5566 TTATCTACCATGATCAAGTTGGCTCATCCTCGGATGCAAGGCTGGTTCAAAATATGCA 5625  
Db 563 euIleHisHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgL 583  
QY 5626 AATCAATAAATGTAGGCCATCACATAAACAGAACCAATGACAAAAACACATGATTATCT 5685  
Db 583 ysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleS 603  
QY 5686 CAATAGATGCAAGAAAGCCCTTTCTCAAAATTCACAGCCCTTCATGCTCAAAAATTCCTCA 5745  
Db 603 erIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuA 623  
QY 5746 GTAACTAGGTATCGATGGATGTATCTCAAAATAAATAAGAGCTATTATTATAC-AAACCCA 5804  
Db 623 snLysLeuGlyIleAspGlyThrTyrLeuLysIleIleArgAlaIleTyrAspLysProt 643  
QY 5805 CAGCAATATCATACTGAATGGGCAAAACTGGAAGCATTCCTCTTCAGAACTGGCACAA 5864  
Db 643 hrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyThrA 663  
QY 5865 GACAAGGATGCCCTCTCTCACCCTCTTCAAGATFACATGGAAGTTCTGGCCAGGG 5924  
Db 663 rgGlnGlyCysProLeuSerProLeuPheAsnIleValLeuGluValLeuAlaArgA 683  
QY 5925 CAATCAGGCATAGAAAGAAATAAGGGTATTCAATAGAAAGAGAGGAAGTCAATTGT 5984  
Db 683 laIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluValLysLeuS 703  
QY 5985 CTCTGTTTGATGATGTTGTATATTATAGAAAAACCCATCGCTCTCAGGCCAAAAAC 6044  
Db 703 erLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnL 723  
QY 6045 TCCTTAAGCTGATAAGCACTTCAGCAAGTCTCAGCAAGTCTCAGCAAGTCTCAGCAAGT 6104  
Db 723 euLeuLysLeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysS 743  
QY 6105 CACAAGACTTCTATATAGCCCAATAATAGACAAACAGAGAGCCCAATCATGAGTGAAC 6164

Db 743 erGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuP 763  
QY 6165 CATTCACAATTCCTACAAAGAGATAAATACCTAGGAATACAACTTACAGGACACAGT 6224  
Db 763 roPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgaspValL 783  
QY 6225 AGGAATCTCTCAAGGAGAACTTACAACCACTGATCAAGGAAATAGAGAGACACAACA 6284  
Db 783 ysAspLeuPheLysGluAsnTyrLysProLeuLeu\*\*\*GluIleLysGluAspThrAsnL 803  
QY 6285 AATGGAAAAACATTCATGCTCAGATAGTAAAGATCAT-----GAAATGCCATAC 6337  
Db 803 ysTrpLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleL 823  
QY 6338 TGCCCAAGATAAATTATAGATTACCTACCTACCTCCCATCAAGCTACCTTCTCTCA 6397  
Db 823 euProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPhePhe 843  
QY 6398 CAGAAATGGAAAAACAACTTTAAATTTTCATATGAAACCAAAAAAGAGCCACAGAGCC 6457  
Db 843 hrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleAla 862  
QY 6458 AAGACAATCTTAGCAAAAAGACAAAGCTGGAGGTATCATGCTACCTGACTTAAACTA 6517  
Db 863 LysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIle\*\*\*LeuProAspPheLysLeu 882  
QY 6518 TACTATTAAGCTACAGTAACCAAACTGCATGGTACTGGTACCAAAAACAGATATATAGAC 6577  
Db 883 TyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsnArgAspIleAsn 902  
QY 6578 CAATGGAAACAGACAGACCTCAGAAATTTACACT-GCAATCTATACCTCTGCTGCTTT 6636  
Db 903 GlnTrpAsnArgThrGluProSerGluIle\*\*\*ProHisIleTyrAsn\*\*\*LeuIlePhe 922  
QY 6637 GACAACTCAGCAAAAACCAAGCAATGAAAAGATTCCCTTATTTAATAATGCTGTGG 6696  
Db 923 AspLysPro\*\*\*LysAsnLys\*\*\*TrpGlyLysAspSerLeuPheAsnLysTrpCysTrp 942  
QY 6697 AAAAATGGCTAGCATATGCAGAAAGCTGAAACTGGATCCCTTCCCTACACCTTATACA 6756  
Db 943 GluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThr 962  
QY 6757 AAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACCCA 6815  
Db 963 LysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeu 982  
QY 6816 GAAGAAACCTTAGCCANTACCATTCAGGATATGCACATGGCAAGACTTCATCACTAAA 6875  
Db 983 GluGluAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPheMet\*\*\*Lys 1002  
QY 6876 ACACCAAAAGCAATGGCAACAAAAGCCAAATAGCAAGTGGGATCTGTATTAACATATAG 6935  
Db 1003 ThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeuLys 1022  
QY 6936 AGCTTCGCACAGCAAAAAAACTGTCATCAGAGTGAACAGCAACCTACAGAAATGGGA 6995  
Db 1023 SerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnProThr\*\*\*TrpGl 1042  
QY 6996 GAAAATTTTGCATCTATCGATCTGACAAAGCTATATATCCAGAGATCTACGAAGAACT 7055  
Db 1042 uLysIlePheAla\*\*\*TyrSerSerAspLysGlyLeuIleSerArgIleTyr\*\*\*GluLe 1062  
QY 7056 TAAACAAATTTACAAGAAAAA---AACACCCCGCTCAAAATATGGCAAGGATATGAG 7111  
Db 1062 uLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAspMetAs 1082  
QY 7112 CAGACACTTCTAAAAGACATATTTATGAGCCCAACAAACATATGAAAAAACCTCATC 7171  
Db 1082 nArgHisPheSerLysGluAspIleTyrAlaAlaLys\*\*\*HisMetLysLysCysSerSe 1102  
QY 7172 ATCATTTGGTGTAGAGAAATGCAAAACCAACACAGTACATACCTATCTGCTAGT 7231  
Db 7231



Db 1102 rSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHisLeuThrProVa 1122  
QY 7232 TAGAATGGTGTACACAAAAGTCAGCAAAACAAATCTGCGAGAGGATGGAGAAAT 7291  
Db 1122 lArgMetAlaIleIleLysSerGlyAsnAsnArgCysTrpArgGlyCysGlyGluI 1142  
QY 7292 AGAACAATTTCCACTGTTGGTGGGAATGTAATTAAGTTCACCAATTTGCGAAGACAGT 7351  
Db 1142 eGlyThrLeuLeuHisCysTrpAspCysLysLeuValGlnProLeuTriPlysSerVa 1162  
QY 7352 GTGGAGATTCCTTAAGGATCTAGACCAAGAATATCATTTGACCCAGCAATCCCATCT 7411  
Db 1162 lTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIleProLeu 1182  
QY 7412 GAGTATATACCAAGGAATATAAATCATCTTATTATAAGACACATGCACACATATGTT 7471  
Db 1182 uGlyIleTyrProLysAspTyrLysSerCysCysTyrLysAspThrCysThrArgMetPh 1202  
QY 7472 TATTGACGACATGATCACAATGCAAGACTTGGGAACCAACCAATGTCATCATGAT 7531  
Db 1202 eIleAlaIleLeuPheThrIleAlaLysThrTrpAsnGlnProLysCysProThrMetI 1222  
QY 7532 AGACTCGATAAGAAACATGGCACATATACACCATGAATACTATGCAGCCATAAAG 7591  
Db 1222 eAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTyrAlaAlaIleLys 1242  
QY 7592 -GATGATTCATGCTCTTTCAGAGATATGGATGAAGCTGGAAACCATCATCTTCAGCAA 7650  
Db 1242 nAspGluPheMetSerPheValGlyThrTrpMetLysLeuGlu\*\*IleIleLeuSerLy 1262  
QY 7651 ACTAACACAGAACAGAAACCAACCAACACATGTTCTCACTGTGAAGTGGAGT 7705  
Db 1262 sLeuSerGln\*\*GlnLysThrLysHisArg\*\*\*PheSerLeuIleGlyGlyAsn 1280

RESULT 3

ABG06053 standard; Protein; 1284 AA.  
XX AC ABG06053;  
DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #6044.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX WO200175067-A2.  
XX PN 11-OCT-2001.  
XX PD 30-MAR-2001; 2001WO-US08631.  
XX PF 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX PA Drmanac RT, Liu C, Tang YT;  
XX PI WPI; 2001-639362/73.  
XX DR N-PSDB; AAS70240.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 20; SEQ ID No 36412; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1284 AA;

Alignment Scores:  
Pred. No.: 0 length: 1284  
Score: 5164.00 Matches: 1049  
Percent Similarity: 78.62% Conservative: 87  
Best Local Similarity: 72.60% Mismatches: 143  
Query Match: 9.87% Indels: 173  
DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x ABG06053 (1-1284)

QY 3401 CGGGCAAAATAACAGCTAGTGTCATAATGGCAGGATCAAAATTCACACATAATAATTA 3460  
Db 1 ArgAlaLysSerProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrLeu 20  
QY 3461 ACCTTAATGTAAATGGCTAAATTCCTCAATTAATAAGACACAGACTGCCAAATGGATA 3520  
Db 21 ThrLeuAsnValAsnGlyLeuAsnSerProLysArgHisArgLeuAlaSerTrpIle 40  
QY 3521 AAGATCAAGACCCATCAGTGTCTGTATTCAGGAGGCCCTCATCATGAAGAAGACACA 3580  
Db 41 LysSerGlnAspProSerValCysCysIleGlnGluThrHisLeuMetCysArgSpThr 60  
QY 3581 CATAGCTCAAAATAAAGGATGGAGAGATTACCAAGTAATGGAACCAACAAAAA 3640  
Db 61 HisArgLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---Lys 79  
QY 3641 AAAGCAGGGGTTGCAATCTCTAGTCTGATAAAACAGACTTTAAACCAACAAGATCAAA 3700  
Db 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99  
QY 3701 AGAGACAAAGAGGCCATTACATAATGTAAAGGCATCAATGAACAAGAGAGCTAAT 3760  
Db 100 ArgAspLysGlyGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThr 119  
QY 3761 ATCCCTAAATATACATGCACCCCAATACAGGAGCCAGATTCATGAAGCAAGTCTTAGA 3820  
Db 120 IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139  
QY 3821 GACCTACAAAGAGACTTTGACTCCCAACAATAATAGTGGGAGTCTAAATAATAATAGA 3880  
Db 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly----- 153  
QY 3881 CACTTTAACACCCCACTGCCAATATATAGCAGATCAATGAGACAGAGAAAAATTACAGGAT 3940  
Db 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173  
QY 3941 ATCCAGGAGTTGAAGTACCTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCC 4000  
Db 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193

```
QY 4001 CACCCCAAAATCAACAGAAATATACACTCTCTCAGACATACATTAACACTATTTTAAAT 4060
Db 194 HisProLysSerThrGluTyrThrPhePheSerAlaProHisHisSerTyrSerLysile 213
QY 4061 GACCATGTAATTTAAAGTAACAACTCCTCAGCAAAATCCAAAAGAACAAATCCTAACA 4120
Db 214 AspHisIleLeuGlySerGluAlaLeuLeuSerLysCysLysArgThrGluIleIleThr 233
QY 4121 AACAGTCTCTCAGACTACAGTGCATCTATTAGAACTCAGAAATTAAGAAATCACTCAA 4180
Db 234 AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln 253
QY 4181 AATCACACAACACTAGTGAACAACTGCTCTGTAAGTACTACTGGGTAAATAAC 4240
Db 254 SerArgSerThrTrpLysLeuAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 273
QY 4241 AAAATGAAGGCAAAATAAGATGTTCTTCAAAACCAATGAGAACAAAGACACAAATGAC 4300
Db 274 GluMetLysAlaGluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyr 293
QY 4301 CAGAACTCTCTGGGCATATTTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCC 4360
Db 294 GlnAsnLeu***AspAlaPheLysAlaValcysArgGlyLysPheIleAlaLeuAsnAla 313
QY 4361 TACAAGAGAAAGCAGGAATATCTAAATAGACACCTTAACATCACAAATTAAGAACTA 4420
Db 314 HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu 333
QY 4421 GAGAAGAAAGCAACAAATTTCAAAAGCTAGCAGAGACAGAAATAACTAAGATCAGA 4480
Db 334 GluLysGlnGluThrHisSerLysAlaSerArgGlnGluIleThrLysIleArg 353
QY 4481 GCAGAACTGAAGAGATAGACACAAAAAGCCCTTCAAAATAAATCAATGAATCCAGGAG 4540
Db 354 AlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSe 373
QY 4541 CTGGTTTTTCAAAAGATCAGCAAAAT----AGACCAGTACAGAGACTAATAAGAGAA 4596
Db 373 rTrpPheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLy 393
QY 4597 AAGAGAGAAGAAATCAAGAGATGCAATAAAAATGATAAGGGGATATCACACCCGATCC 4656
Db 393 sArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrAspPr 413
QY 4657 CACAGAAATACAACATATTATCAGAGAAATTTATAACACCTCTATGCAATAAACAATCA 4716
Db 413 oThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGl 433
QY 4717 AAATCTAGAAGAAATGGATAAATTCCTGGACACATATGTAGCCCTGTATGGACCTTGGGG 4776
Db 433 uAsnLeuGluGluMetAspThrPheLeuAsp----- 443
QY 4777 ACAGAACAAAAGGGGTGAATGCAGAAATAAAGACAAAGACAAAAGATGTTTGGAA 4836
Db 443 ----- 443
QY 4837 GTAGGGGTACGGGGCAACTTGGCTCTAATGCAGACGGCCCTGAGCTTTACACCACCT 4896
Db 443 ----- 443
QY 4897 CTGTATTATTAGCAAAAGAGATAGCGAGAGGGTGAGTTGGAAGAGAGGTACAGCTGTT 4956
Db 443 ----- 443
QY 4957 AGGTCCAGAGTAGGCCTGCAAGACTGCATTCCTCAACAATAGCTCTAGATGTCCAGT 5016
Db 443 ----- 443
QY 5017 AGATAACCTCAAGGAGCCAGTGCCAGGGAGTGATGGCCCTCAGCAAAACCTTCTAGGCGAG 5076
Db 443 ----- 443
QY 5077 GCACAGAGTAAGTTTCCACATCTCTGATTACGATAAACAGTTTGGCTTGTGATCAA 5136
```

---

```
Db 443 ----- 443
QY 5137 GTACCTCCAGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAA 5196
Db 443 ----- 443
QY 5197 CACATACACCTCTCAAGACTAAACACAGGAAGTCAAAATCCCTGAATATACAGTAAC 5256
Db 444 -ThrTyrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleTh 463
QY 5257 AAGTTCCTAAATTCAGCAGCAATATTGATAGCTTACCACCAACCAAAAAAGTCCAGACACAGA 5316
Db 463 rGlyAlaGluIleValAlaIleIleAsnSerLeuSerThrLysLysSerProGlyProAs 483
QY 5317 CGGATTCCACAGCCAAATCTACCAGAGGTACAAAGAGAAGCTGTACTATTCTCTCTGAA 5376
Db 483 pGlyPheThrAlaGluPheTyrGlnArgTyrLysGluGluLeuValProPhePheLeuLy 503
QY 5377 ACTATTCAAAAAATAGAA---AATGGAAATCTCCCTAACTCATTTTACAGAGCCACGA 5433
Db 503 sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI 523
QY 5434 TCATCTGTATACCAAAACCTTAGCAGTGCACACAAAAAGAGGAATTTTCAGGCCCATAT 5493
Db 523 leIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS 543
QY 5494 CCCTGATGAACATTGTGAAATCCTCAATAAATACTGGCAAAACCAAAATCCAGCAGC 5553
Db 543 erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGlnH 563
QY 5554 ACATCAAAAAAGCTTATCTACCATGATCAAGTGGCGTCATCCCTGGATGCAAGCGTGGT 5613
Db 563 isIleLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTrp 583
QY 5614 TCAAAATATGCAATCAATAAATGTAGCCATCATATAACAGAACCAATGACAAAAACC 5673
Db 583 heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH 603
QY 5674 ACATGATATTCTCAATAGATGCAGAAAGCGCTTTGTCAAAATTTCAACAGCCCTTCATGC 5733
Db 603 sMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL 623
QY 5734 TAAAAATTCAGTAATAACTAGGTATCGATGGAATGTATCTCAAAATTAATAGAGCTATTT 5793
Db 623 euLysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT 643
QY 5794 ATAC-AAACCCACAGCAATATATCTGATGGGCAAAACTGGAAGCATTCCTCTTTGA 5852
Db 643 yrrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL 663
QY 5853 GAATGGCACAGCAAGAGTGGCTCTCTCACCACCTCCTATTCAGATATACTATTGGAAG 5912
Db 663 ystThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP 683
QY 5913 TTTCTGGCCAGGCAATCAGGCATAGAAAGAAATAAAGGTATTCAATAGAAAGAGAGG 5972
Db 683 heLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluG 703
QY 5973 AAGTCATATGCTCTCTGTTTGCAGATGACATGTTTGTATATTTTAAAAACCCCTCGTCT 6032
Db 703 luValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS 723
QY 6033 CAGGCCCAAAATCTCTTAAGCTGATAAGCAACTTCAGCAAAAGTCTCAGGACACAAAAATCA 6092
Db 723 erAlaGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrLysIleA 743
QY 6093 ATGTGCAAAATCACAAGCATTTCTTATCGCCCAATAATAGACAAACAGAGAGCCAAATCA 6152
Db 743 snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM 763
QY 6153 TGAGTGAACCTCATTCACAAATTCCTACAAAGAGAATAAATACCTAGGATACAACTTA 6212
```

Qy	7280	ATGTGGAGAAATAGGAACACATTTTCCACTGTTGGTGGGAATGTAAATTAGTTCAACCACTT	7339
Db	1142	YCYSGLYGLUIEGLYThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLe	1162
Qy	7340	GTGGAACACAGTGTGGAGATTCCCTTAAGGATCTAGAACACAGAAATATCATTTGACCCAGC	7399
Db	1162	uTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAl	1182
Qy	7400	AATCCCATTTACTGAGTATATACCCAAAGGAATATAAATCATTTCTATTTATAAGACACATG	7459
Db	1182	alleProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysCysTyrLysAspThrCy	1202
Qy	7460	CACACATATGTTTATTCACACACTGATCACAATAGCAAGAACTGGGAACCAACCCAAATG	7519
Db	1202	sThrArgMetPheIleAlaIleLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy	1222
Qy	7520	TCCATCAGTGATAGACATGGATAAAGAAAACATGCACATATACACCATGAATAACTATGC	7579
Db	1222	sProThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTrAl	1242
Qy	7580	AGCCATAAAAG-GATGAGTTCATGTCCTTTGCAGAGATATGGATGAAGCTGGAACCCAT	7638
Db	1242	aAlaIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIl	1262
Qy	7639	CATTCTCAGCAAACTACACAGAACAACAGAAAACCAACACACATGTTCTCACTTGTAAAG	7698
Db	1262	eIleLeuSerLysLeuSerGlnGluGlnLysThrLysHisArgIlePheSerLeuIleGl	1282
Qy	7699	TGGGAGT 7705	
Db	1282	yGlyAsn 1284	
RESULT 4			
ABG09636			
ID	ABG09636	standard; Protein; 1284 AA.	
XX	AC	ABG09636;	
XX	DT	13-FEB-2002 (first entry)	
XX	DE	Novel human diagnostic protein #9627.	
KW	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.	
OS	OS	Homo sapiens.	
XX	FN	WO200175067-A2.	
XX	PD	11-OCT-2001.	
XX	PF	30-MAR-2001; 2001WO-US08631.	
XX	XX	31-MAR-2000; 2000US-0540217.	
PR	PR	23-AUG-2000; 2000US-0649167.	
XX	XX	(HYSE-) HYSEQ INC.	
PA	PA	Drmanac RT, Liu C, Tang YT;	
PI	PI	WPI; 2001-639362/73.	
XX	XX	N-PSDB; AAS73823.	
XX	XX	New isolated polynucleotide and encoded polypeptides, useful in	
PT	PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	PT	responsible for genetic disorders or other traits and to assess	
PT	PT	biodiversity -	
XX	XX	Claim 20; SEQ ID NO 39995; 103pp: English.	
PS	PS	The invention relates to isolated polynucleotide (I) and	
XX	XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	CC		

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 1284 AA;

# Alignment Scores:

Pred. No.: 0 Length: 1284  
 Score: 5164.00 Matches: 1049  
 Percent Similarity: 78.62% Conservative: 87  
 Best Local Similarity: 72.60% Mismatches: 143  
 Query Match: 9.87% Indels: 173  
 DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x ABG09636 (1-1284)

QY 3401 CGGGCAAAATACACAGTAGTGTCATAATGGCAGGATCAAAATTCACACATAATAATTA 3460  
 DB 1 ArgAlaLysSerProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrIleLeu 20  
 QY 3461 ACCTTAATGTAAATGGCTAAATTCGCCAATTAAGACACAGACTGGCAATTTGGATA 3520  
 DB 21 ThrLeuAsnValAsnGlyLeuAsnSerProIleTyrArgHisArgLeuAlaSerTrpIle 40  
 QY 3521 AAGACTCAAGACCCATCAGTGTCTGTTATTCAGGAGGCCATCTCACATGAAGAACACACA 3580  
 DB 41 LysSerGlnAspProSerValCysCysIleGlnGluThrHisLeuMetCysArgAspThr 60  
 QY 3581 CATAGGCTCAAAATAAGGGATGGAGGAGATTTACCAAGTAATGGAACCAACAAAAA 3640  
 DB 61 HisArgLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln--Lys 79  
 QY 3641 AAGCAGGGTTCGAATCCTAGTCTCTGATAAACAGACTTTAAACCAACAAAGATCAA 3700  
 DB 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99  
 QY 3701 AGAGCAAGAGGCGCATACATATGTAAGGCATCAATGGACAGAGAGAGCTAACT 3760  
 DB 100 ArgAspLysGluGlyHisTyrMetMetValLysGlySerIleGlnGluLeuThr 119  
 QY 3761 ATCTTAATATACATGCCCAATACAGAGCACCCAGATTCATAAAGCAAGTCTTTAGA 3820  
 DB 120 IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139  
 QY 3821 GACCTACAAAGAGACTTTGACTCCACACAATATAGTGGGAGTCTTAATAATAATAGA 3880  
 DB 140 AspLeuGlnArgAspLeuAspSerHisThrIleMetGly----- 153  
 QY 3881 CACTTTAACCCCACTGCCAATATAGGCGAGATCAATGAGACAGAAAAATTAACAAGAT 3940  
 DB 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173  
 QY 3941 ATCCAGGAGTTGAATGAGCTGGACCAAGCGGACCTTAATAGATATCTACAGAACTCCC 4000  
 DB 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193  
 QY 4001 CACCCCAATCAACAGATATACACTCTCTCAGCATCATATACACCTATTTAAAATT 4060

DB 194 HisProLysSerThrGluTyrThrPhePheSerAlaProHisHisSerTyrSerLysIle 213  
 QY 4061 GACCATGTAAATTTAAGTAACAACACTCCTCAGCAAAATGCAAAAGAACAAATCCCTAACA 4120  
 DB 214 AspHisIleLeuLysGlySerGlnAlaLeuLeuSerLysCysLysArgThrGluIleIleThr 233  
 QY 4121 AACAGTCTCTCAGACTACAGTACGCAATCTATTAGAACTCAGAAATTAAGAACTCACTCAA 4180  
 DB 234 AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln 253  
 QY 4181 AATCACAACTACATGGAACACTGAACACCTGCTCTGAATGACTACTGGGTAATAATAC 4240  
 DB 254 SerArgSerThrThrTrpLysLeuAsnAsnLeuLeuLeuLeuAsnAspTyrTrpValHisAsn 273  
 QY 4241 AAAATGAAGGCAAAATAAAGATGTCTTTGAAACCAATGAGAACAAAGACACACATGTAC 4300  
 DB 274 GluMetLysAlaGluIleLysMetPheGluThrAsnGluAsnLysAspThrThrTyr 293  
 QY 4301 CAGAACTCTGGGGCATATTTAAAGCAGTGTAGAGGGAATTTATAGCACTAGATGCC 4360  
 DB 294 GlnAsnLeu\*\*\*AspAlaPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAla 313  
 QY 4361 TACAAGAAAGCAGGAAATATCTAAATAGACACCTTAACATCAATTAAGAACTA 4420  
 DB 314 HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu 333  
 QY 4421 GAGAAGAAAGACGCAAAATTTCAAAAGCTAGCAGACAGCAAGAAATTAACTAAGATCAGA 4480  
 DB 334 GluLysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArg 353  
 QY 4481 GCAGAACTGAAGGAGATAGACACACAAAAGCCCTTCAATTAATCAATGAATCAGGAG 4540  
 DB 354 AlaGluLeuLysGluIleGluThrGlnLysThrLysThrLeuGlnLys-IleAsnGluSerArgSe 373  
 QY 4541 CTGGTTTTTGAAGAAGATCAGCAAAAT-----AGACCCTAGACACTAAATAAGAAAGAA 4596  
 DB 373 rTrpPheGluArgIleAsnLysIleAspArgProLeuAlaArgLeuIleLysLysLys 393  
 QY 4597 AAGAGAAAGATCAAGAGATGCAATATAAAGGATATAAAGGGATATCACCACCGATCC 4656  
 DB 393 sArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspPr 413  
 QY 4657 CACAGAAATACAACTATTATCAGAGAATATTATAAACACCTCTATGCAATAAATACTAGA 4716  
 DB 413 oThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGln 433  
 QY 4717 AAATCTAGAAATGGATAAATTCCTCGACACATATGTAGCCTGTATGGACCTTGGGGG 4776  
 DB 433 uAsnLeuGluMetAspThrPheLeuAsp----- 443  
 QY 4777 ACAGAACAAAAGGGGTGAATGCAGAAATAAAGACAAAGAGAGATGTGTTGCA 4836  
 DB 443 ----- 443  
 QY 4837 GTAGGGGTGAGGGGCAACTTGCTCTAATGGCAAGGGCCCTGAGCTTTACACCACCT 4896  
 DB 443 ----- 443  
 QY 4897 CTGTATTTATTAGCAAAAGAGATAGCGAGAGGTGAGTGGAGAAGAGAGGTGCTGTT 4956  
 DB 443 ----- 443  
 QY 4957 AGGTCCAGAGTAGGCTGCAAGACTGCATTCTCTCAACAATAGGCTCTAGATGTCCCACT 5016  
 DB 443 ----- 443  
 QY 5017 AGATAACCTCAAGGAGCGAGTGCCAGGGAGTGATGGCCCTCAGCAAACTTTCTAGGGCAG 5076  
 DB 443 ----- 443  
 QY 5077 GCACAGAAGTAAGTTTGGCCACATCTCTATTATTCAGATAAACAGTTTCTGCTGTTGATCAA 5136

Db	443	-----	443
Qy	5137	GTAGCCTCCAGTGAATGCTGAGTTGGTGCATGATCCCTTTGGCCTTTGGCTCCCAAAA	5196
Db	443	-----	443
Qy	5197	CACATACACCTCTCAAGACTAAACACGAGAGTCAAAATCCCTGTAATATACACAGTAAC	5256
Db	444	-ThrTyr-ThrLeuProArgLeuAsnGlnGluGluValGluSerLeuAsnArgProIleTh	463
Qy	5257	AAGTTCTCAAAATCAACGACGTAATGATAGCCCTACCAACCAAAAAAGTCCAGGACAGA	5316
Db	463	rglyAlaGluValAlaIleLeuAsnSerLeuSerThrLysLysSerProGlyProAs	483
Qy	5317	CGGATTCACAGCCAAATTCACACAGAGTCAAGAGAGAGTGTACTATTCTCTCTCGAA	5376
Db	483	pglyPheThrAlaGluPheTyrGlnArgTyrLysGluGluLeuValProPhePheLeuLy	503
Qy	5377	ACTATTCCTCAAAAAATAGAA---AATGGGAATCCCTCACTCACTCACTTTTACGAGCCAGCA	5433
Db	503	sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI	523
Qy	5434	TCATCCGTGATACCAAAACCTAGCAGTGCACACACAAACAAAGAGGAAATTTGAGGCCATAT	5493
Db	523	IleIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS	543
Qy	5494	CCCTGATGAACATGTGATGCTGAAATCCCTCAATAAATPACTGGCAAAACCAATCCAGCAGC	5553
Db	543	erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnGlnH	563
Qy	5554	ACATCAAAAAGCTTATACATGATCAAGTTGGCGTCACTCCCTGGGATCCAAAGGCTGTT	5613
Db	563	isIleLysLysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnGlyTrp	583
Qy	5614	TCRAAATATCAATCAATCAATAGTACGATCAAGTTGGCGTCACTCCCTGGGATCCAAAGGCTG	5673
Db	583	heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH	603
Qy	5674	ACATGATATCTCAATAGATGCAGAAAAGGCTTTGTCAAAATCAACAGCCCTTCATGC	5733
Db	603	isMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL	623
Qy	5734	TAAAAATTCAGTAAGTATAGTATCGATGGAATGATATCAAAATATTAAGAGCTATTT	5793
Db	623	eulysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT	643
Qy	5794	ATAC-AAACCCACAGCAATATCATCTGATGAGTGGCAAAACCTGGAAGCATTCCTTTGA	5852
Db	643	yrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL	663
Qy	5853	GAACTGGCACAAAGGATGGCTCTCTCACACCTCTCTTATTCAGATACTATTGGAAG	5912
Db	663	ysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP	683
Qy	5913	TTCTGGCCAGGCAATCAGGCATAGAAAGAAATTAAGGTATTCAAATAGAAAGAGG	5972
Db	683	heLeuAlaArgAlaIleArgGlnGluLysGluIleGlnLysGlyIleGlnLysGluG	703
Qy	5973	AAGTCATATGCTCTCTTTGACAGATGACATGTTGTATATTAGAAACCCCATCGCT	6032
Db	703	IuValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS	723
Qy	6033	CAGGCCAAAACCTCTTAAGCTGATAGCAACTTCAGCAAGTCTCAGGACACAAAATCA	6092
Db	723	erAlaGlnAsnLeuLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrLysIleA	743
Qy	6093	ATGTGCAAAATCACAAGCTTCTTATACGCCCAATATAGCAAAACAGAGAGCCAAATCA	6152
Db	743	snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM	763
Qy	6153	TGAGTGACCTCTATTACAAATTTGCTACAAGAGATAAATACCTAGGAATACAACTTA	6212
Db	763	etSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuT	783
Qy	6213	CAAGGGACAGCTAGGAACTCTTCAAGGAGAACTACAAACCACTGATCAAGGAAATAAGAG	6272
Db	783	hrArgspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluLeuLysG	803
Qy	6273	AGGACACAAACAAATGAAACCAATTCATGCTCACAGATAGTAGAATCAT-----G	6325
Db	803	IuAspThrLysLysTrpLysAsnThrProCysSerTrpValGlyArgIleAsnIleValL	823
Qy	6326	AAATGCCATCTACCCCAAGTAAATATAGATTCAGTACCTACCCCATCAAGTACCAT	6385
Db	823	ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProM	843
Qy	6386	TGACTTTCTTCACAGAAATTCGAAAAACAACTTTTAAATTTTCATATGCAACCAAAAAAGA	6445
Db	843	etThrPheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArg	862
Qy	6446	GCCACAGACCCAGACAACTTAAAGCAAAAGAACAAAGCTGGAGGTATCATGCTACCT	6505
Db	863	AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro	882
Qy	6506	GACTTAAACTATACTATAAGGCTACAGTACCAACCAACTGCATGCTACTGTACCAAAAC	6565
Db	883	AspPheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTrpTyrGlnAsn	902
Qy	6566	AGATATATAGACCAATGGAACAGAACAGACCTTCAGAAATTTACACT-GCAATCTACATC	6624
Db	903	ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn	922
Qy	6625	CATCTGATCTTGCACAACTGCACAAACAGCAATGGAAGAGTTCCTTATTAAT	6684
Db	923	TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn	942
Qy	6685	AAATGGTGTGGAAAACTGCTAGCCATATGCAGAAAGCTGAAACCTGCTCCCTCCTT	6744
Db	943	LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValAspProPheLeu	962
Qy	6745	ACACCTTATACAAAGTTAACTCAAGATGAATTAAGACTTAATATAAGACATATAAAC	6804
Db	963	ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr	982
Qy	6805	ATAAAACCCA-GAAGAAAACCTAGGCAATACCATTCAGATATGGACATGGGCAAGAC	6863
Db	983	IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysAsp	1002
Qy	6864	TTCATGACTAAAACACCAAGCAATGGCAACAAACCCCAAAATAGACAGTGGAGTCTG	6923
Db	1003	PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu	1022
Qy	6924	ATTAACATATAGAGCTTCTGCACAGCAAAAAAACTGTCATCAGAGTGAACAGCAACC	6983
Db	1023	IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr	1042
Qy	6984	TACAGATGGGAGAAATTTTTCGATCTCATGATCTGCAAAAGGCTTAATATCCAGAGAT	7043
Db	1042	oThrThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIl	1062
Qy	7044	CTAGGAAGACTTAAACAAATTTTACAGAAAAA- - -AACACCCCTGCAAAATATCGGC	7099
Db	1062	eTyrAsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAl	1082
Qy	7100	AAAGGATATAGGACACACTTCTCAAAAGAGACATTTATGCAAGCAACCAATATGAA	7159
Db	1082	aLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy	1102
Qy	7160	AAAAACCTCATCATCTATTGCTGCTTAGAGAAATGCAAAACAAACACACATGACATCA	7219
Db	1102	sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHi	1122
Qy	7220	TCATCATCTAGTGTAGATGCTGATCACTAAAGACTCAGAAACAAACAAATGCTGGAGG	7279
Db	1122	sLeuThrProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGl	1142

QY 7280 ATGTGGAGAAATAGGAACACTTTCCACTGTTGGGGAATCAATTAAGTTCAACCAT 7339  
 Db 1142 yCyseGlyGluIleGlyThrLeuLeuHisCysTrpAspCysLysLeuValGlnProLe 1162  
 QY 7340 GTGGAAGACAGTGTGGAGATTCCTTAAGGATCTAGAACCCAGAAATATCATTTGACCCAGC 7399  
 Db 1162 utrPlysservaltPargPheLeuArgAspLeuGluLeuGluIleProPheAspProAl 1182  
 QY 7400 AATCCCATTTACTGATATATACCCAAAGGAATATAAATCATCTCTATTATAAGACACATG 7459  
 Db 1182 alileProLeuLeuGlyIleTyrProAsnAspTyrLysSerCysTyrLysAspThCy 1202  
 QY 7460 CACACATATGTTTATGAGCACTGATCACAATAGCAAGACTTGAACCAACCCAAATG 7519  
 Db 1202 sThrArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy 1222  
 QY 7520 TCATCAGTGTAGACTGGATAAGAAACATGCGACATATACACCATGAATATATATGC 7579  
 Db 1222 sProThrMetIleAspTrpIleLysLysMetTrpHisIleTyrThrMetGluTyrTrAl 1242  
 QY 7580 AGCCATAAAAG-GATGAGTTTCATGCTCTTTCGACAGATATGGATGAAGCTGGAACCAT 7638  
 Db 1242 aaIalleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrIl 1262  
 QY 7639 CATTCAGCAACTAATACACAGAACAGAAACCAACCAACCACTGTTCTACTGTGAAG 7698  
 Db 1262 eileLeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGl 1282  
 QY 7699 TGGGAGT 7705  
 Db 1282 yGlyAsn 1284

RESULT 5  
 ABG10795  
 ID ABG10795 standard; Protein; 1284 AA.  
 AC ABG10795;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #10786.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS74982.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 41154; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful for medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1284 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 1284  
 Score: 5164.00 Matches: 1049  
 Percent Similarity: 78.62% Conservative: 87  
 Best Local Similarity: 72.60% Mismatches: 143  
 Query Match: 9.87% Indels: 173  
 DB: 22 Gaps: 7

US-10-083-853b-2 (1-29921) x ABG10795 (1-1284)

QY 3401 CGGCAAAATAACCACTAGTGTCTATAATGCGAGGATCAATTCACATATAATATTA 3460  
 Db 1 ArgAlaLysSerProAlaAsnIleIleMetThrGlySerAsnSerHisIleThrLeu 20  
 QY 3461 ACCTTAATGTAAATGGCTAAATTCCTCCCAATTAAGACACACAGCTGGCAATTTGGATA 3520  
 Db 21 ThrLeuAsnValAsnGlyLeuAsnSerProIleLysArgHisArgLeuAlaSerTrpIle 40  
 QY 3521 AAGAGTCAAGACCCATCATGCTGTCTATTTCAGGAGGCCCTCTCACATGAACACACA 3580  
 Db 41 LysSerGlnAspProSerValCysIleGlnGluThrHisLeuMetCysArgAspThr 60  
 QY 3581 CATAGGCTCAAAATAAAGGATGGAGGAAGATTTACCAAGTAATTTGGAACCAACAAAAA 3640  
 Db 61 HisArgLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---Lys 79  
 QY 3641 AAAGCAGGGTTCGAATCCTAGTCTCTGATATAAACACAGACTTTAAACCAACAAAGATCAA 3700  
 Db 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99  
 QY 3701 AGAGCAAAAGAGGCCATTACATATGCTAAAGGCATCAATGGCAACAAAGAGCTAACT 3760  
 Db 100 ArgAspLysGluGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThr 119  
 QY 3761 ATCTTAATATACATGACCCCAATACAGGAGCACCAGATTCATAAAGCAAGTTCTTTAGA 3820  
 Db 120 IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139  
 QY 3821 GACCTACAAGAGACTTTGACTCCACACATATATAGTGGGAGTCTAAATATAATAATAGA 3880  
 Db 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly----- 153  
 QY 3881 CACTTTTAACACCCCACTGCCAATATTAGGCGACATCAATGACGACAGAAAAATTAACAAGGAT 3940  
 Db 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173  
 QY 3941 ATCCAGGAGTTGAAGTCTGAGCTCTGGACCAAGCGGACCTTAATAGATATCTACAGAACTCCC 4000  
 Db 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193  
 QY 4001 CACCCCAAAATCAACAGATATATACACTCTTCTCAGCATCACATTACACCTTTTAAATTT 4060  
 Db 174 ThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193









polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG3077 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 1284 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 1284  
 Score: 5164.00 Matches: 1049  
 Percent Similarity: 78.62% Conservatives: 87  
 Best Local Similarity: 72.60% Mismatches: 143  
 Query Match: 9.87% Indels: 173  
 DB: 22 Gaps: 7

US-10-083-853B-2 (1-29921) x ABG14889 (1-1284)

3401 CGGCAAAATAACAGCTAGTGTCTAATGCGAGGATCAATTCACACATAATAATTA 3460  
 1 ArgAlaLysSerProAlaAsnIleMetThrGlySerAsnSerHisIleThrLeu 20  
 3461 ACCTTAATGTAATGGCTAAATCCCAATTAAGACACACAGCTGGCAATGGATA 3520  
 21 ThrLeuAsnValAsnGlyLeuAsnSerProIleLysArgHisArgLeuAlaSerTrpIle 40  
 3521 AAGAGTCAGACCCATCAGTGTGTCTATTGAGGAGGCCATCTCACATGAAAGACACA 3580  
 41 LysSerGlnAspProSerValCysCysIleGlnGluThrHisLeuMetCysArgAspThr 60  
 3581 CATAGGCTCAAAATAAGGATGAGGAGGATTTACCAAGTAATGGAACAAAAA 3640  
 61 HisArgLeuLysIleLysGlyTrpArgLysIleTrpGlnAlaAsnGlyLysGln---Lys 79  
 3641 AAAGCAGGGTGTCAATCTCTGCTGATAAAGACAGCTTTAAACCAACAAAGATCAA 3700  
 80 LysAlaGlyValAlaIleLeuValSerAspLysThrAspPheLysProThrLysIleLys 99  
 3701 AGAGACAAGAGGCCATTACATAATGTAAGGCATCAATGGAACAAGAGAGCTAACT 3760  
 100 ArgAspLysGluGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThr 119  
 3761 ATCTTAATATACATGACCCATACAGGAGCACCAGATTCATTAAGCAAGTTCTTGA 3820  
 120 IleLeuAsnMetTyrAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuSer 139  
 3821 GACCTACAAAGAGACTTGTACTCCACACAATAATAGTGGGAGCTACTAATAATAATAGA 3880  
 140 AspLeuGlnArgAspLeuAspSerHisThrLeuIleMetGly----- 153  
 3881 CACTTTAACCCCTGCGCAATATTAGGCAGATCAATGAGACAGAAAAATTAACAAGGAT 3940  
 154 AspPheAsnThrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAsp 173  
 3941 ATCCAGAGTGTGAAGTCTGAGCCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4000  
 174 ThrGlnGlnLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeu 193  
 4001 CACCCCAAAATCAACAGAAATATACACTCTCTCTGAGCATCATATACACCTATTTAAAT 4060  
 194 HisProLysSerThrGluTyrThrPhePheSerAlaProHisSerTyrSerLysIle 213

4061 GACCATGTAATTTAAGTAAACACCTCTCAGCAATGCAAAAGACAGAAATCCCTAAACA 4120  
 214 AspHisIleLeuGlySerGluAlaLeuLeuSerLysCysLysArgThrGluIleIleThr 233  
 4121 AACAGCTCTCAGACTACAGTCAATCTATTAGAACTCAGAATTAAGAAACTCAGTCAA 4180  
 234 AsnTyrLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysAsnLeuThrGln 253  
 4181 AATCACAACTACATGGAACCTGCAACCTCTCTGATGACTACTGGTAAATAAC 4240  
 254 SerArgSerThrThrTrpLysLeuAsnLeuLeuAsnAspTyrTrpValHisAsn 273  
 4241 AAAATGAAGCAAAATAAAGATGTTCTTTGAACCAATGAGAACCAAGACACAATGTAC 4300  
 274 GluMetLysAlaGluIleLysMetPhePheGluThrAsnGluAsnLysAspThrThrTyr 293  
 4301 CAGAACTCTGCGGGCATATTAAAGCAGTGTAGAGGGAAATTTATAGCACTAGATGCC 4360  
 294 GlnAsnLeu\*\*\*AspAlaPheLysAlaValCysArgGlyLysPheIleAlaLeuAsnAla 313  
 4361 TACAAGAGAAAGCAGGAAATATCTAAATAGACACCTTACATCACAAATTAAGAACTA 4420  
 314 HisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeu 333  
 4421 GAGAAGAAGAGCAAAATAATTCAAAGCTTAGCAGAGACAGCAAGAAATACTAAGATCAGA 4480  
 334 GluLysGlnGluGlnThrHisSerLysAlaSerArgGlnGlnIleThrLysIleArg 353  
 4481 GCAGAACTGAAGGATAGACACAAAAGCCCTCAATAATAATCAATGAATCCAGGAG 4540  
 354 AlaGluLeuLysGluIleGluThrGlnLysThrLysLysIleAsnGluSerArgse 373  
 4541 CTGGTCTTTGAAGATCAGCAAAAT---AGACCACCTAGACAGACTATTAAGAAAGAA 4596  
 373 TrpPhePheGluArgIleAsnLysIleAspArgProLeuAlaLeuIleLysLysLys 393  
 4597 AAGAGAGAAGAAATCAAGAGATGCAATATAAATAAAGGGGATATACACCCAGTCC 4656  
 393 ArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrAspPr 413  
 4657 CACAGAAATACAACTATTATCAGAGAAATTAATAACACCTCTATGCAAAATAAATAGA 4716  
 413 OThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuG 433  
 4717 AATCTAGAGAAATGATTAATCTCGGACACATATGTAGCCTGTATGGACCTTGGGG 4776  
 433 uAsnLeuGluGluMetAspThrPheLeuAsp----- 443  
 4777 ACAGAAACAAAGGGGTGAATGCGAAATAAAGACAAACAGAGATGATCTTTGGAA 4836  
 443 ----- 443  
 4837 GTAGGGTCAGGGGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCACCT 4896  
 443 ----- 443  
 4897 CTCTATTATTAGGCAAAAGAGATACGAGAGGGTGTGGTGGGAAGAGGTCAGCTGTT 4956  
 443 ----- 443  
 4957 AGGTCCAGAGTAGGCGCTGCAAGACTGCATTCTCAAAACAATAGGCTCTAGATGCCAGT 5016  
 443 ----- 443  
 5017 AGATAACCTCAAGGAGCCAGTGCCAGGAGTGATGCCCTCAGCAAAACCTTCTTAGGGCAG 5076  
 443 ----- 443  
 5077 GCACAGAAGTAAGTTTGGCCACATTTCTGTTATTCACGATAAAGTTGCTGTTGATCAA 5136  
 443 ----- 443

```
QY 5137 GTAGCCTCAGTGGAAATGCTGAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAA 5196
Db 443 ----- 443
QY 5197 CACATACACCCTCTCAAGACTAAACACGAGAAAGTCAAAATCCCTGAATATACCAAGTAAC 5256
Db 444 -ThrTyrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleTh 463
QY 5257 AAGTTCTAAATTAAGCAGCAATATGATAGCTTACCAACCAAAAAAGTCCAGGACCAAGA 5316
Db 463 rGlyAlaGluIleValAlaIleAsnSerLeuSerThrLysLysSerProGlyProAs 483
QY 5317 CGGATTACAGCCAAATTTCTACAGAGGTACAAAGAGAGTGGTACTATTCCTTCTGAA 5376
Db 483 pGlyPheThrAlaGluPheTyrGlnArgTyrLysGluLeuValProPhePheLeuL 503
QY 5377 ACTATTCCAAAAATAGAA---AATGGGAATCCCTTAACCTCATTTTACGAGGCCAGCA 5433
Db 503 sLeuPheGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrGluAlaSerI 523
QY 5434 TCATCCTGATACCAAACTAGCAGTGACACAAACAAAGAGGAAATTTACAGGCCATAT 5493
Db 523 leIleLeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleS 543
QY 5494 CCCTGATGAACATGTGTAATCTCAATATAAATCTGGCAAAACCAATCCAGCAGC 5553
Db 543 erLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnGlnIleGlnH 563
QY 5554 ACATCAAAAGCTTATCTACCATGATCAAGTTGGCTGTCATCCCTGGGATGCAAGCTGTT 5613
Db 563 isIleLysLysLeuIleHisHisaspGlnValGlyPheIleProGlyMetGlnGlyTrpp 583
QY 5614 TCAAAATATGCAATCAATAATATGAGCCATCATATAAACAGAACCAATGACAAAAACC 5673
Db 583 heAsnIleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnH 603
QY 5674 ACATGATTATCTCAATAGATCCAGAAAGGCTTTGTCAAAATTCACAGCCCTTCATGC 5733
Db 603 isMetIleValSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetL 623
QY 5734 TAAAAATTTCTCAGTAAACTAGGTATCGATGGAATGTATCTCAAAATATAAGAGCTATTT 5793
Db 623 euLysThrIleAsnLysIleGlyIleAspGlyMetTyrPheLysIleIleArgAlaIleT 643
QY 5794 ATAC-AAACCCACAGCCAAATATCATATCTAGTGAATGGCAAAACTGGAAGCATTCCTTTGA 5852
Db 643 yrAspLysThrThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuL 663
QY 5853 GAACCTGGCACAAGACAGGATGCCCTCTCTCACCACCTCTATTCAAGNACTATATGGNAG 5912
Db 663 ysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluP 683
QY 5913 TTCTGGCCAGGCAATCAGGCAATAGAAAGAAATAAGGGTATTCAATAGAAAGAGAGG 5972
Db 683 heLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluG 703
QY 5973 AAGTCATATCTCTCTGTTTCGACATGACATGTTTGTATATTTTGAACCAACCCCTCGTCT 6032
Db 703 luValLysLeuSerLeuPheAlaAspMetIleValAspLeuGluAsnProIleValS 723
QY 6033 CAGGCCAAAACTCCTTAAGCTGATAGCAACTTCAGCAAGTCTCAGGACACAAAATCA 6092
Db 723 erAlaGlnAsnLeuLysLeuIleSerAsnPheSerLysLeuSerGlyTyrIleA 743
QY 6093 ATGTGCAAAATATCAAGCATCTTATAGCGCAATAATAGACAAACAGAGAGCCAAATCA 6152
Db 743 snValGlnLysSerGlnAlaPheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleM 763
QY 6153 TGAGTGAACCTTCATTCACATTCCTACAAAGAGAATAAAATACCTAGGAATACAACTTA 6212
Db 763 etSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuT 783
QY 6213 CAAGGGACACGTAGGAATCTTCAAGGAGAACTACAAACCCACTGATCAAGGAAATAAGAG 6272
```

---

```
783 hrArgAspValLysAspLeuPheLysGluAsnTyrLysProLeuLeuLysGluIleLysG 803
6273 AGGACACAAACAAATGGAACAAATTCATCATCTCACAGATAGTAAGATCAT- ----G 6325
803 luAspThrLysLysTrpLysAsnThrProCysSerTrpValGlyArgIleAsnIleValL 823
6326 AAAATGCCATACTGCCAAAGTAAATATATAGATTACAGTGTACCCCATCAAGCTACCAT 6385
823 ysMetAlaIleLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProM 843
6386 TGACTTTCTTCCACAGAAATGGAACAAACAACTTTAAATTTTCATATGTAGAACCAAAAAGA 6445
843 etThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArg 862
6446 GCCCACAGAGCCAAAGACAATCTTAAGCAAAAGAACAAAGCTGAGGTATCATCTGCTACT 6505
863 AlaArgIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuPro 882
6506 GACTTAAACCTATATTAAGGCTACAGTAAACCAAACTGCATGCTACTGGTACCAAAAC 6565
883 AspPheLysLeuTyrTyrLysAlaThrValThrLysThrAlaTrpTyrTriTyrGlnAsn 902
6566 AGATATATAGACCAATGGAACAGACAGACCTCAGAAATTACACT-GCAATCTACATC 6624
903 ArgAspIleAspGlnTrpAsnArgThrGluProSerGluIleThrProLeuThrTyrAsn 922
6625 CATCTGATCTTTGACAACTCGACAAACAAACAAAGCAATGGAAAAAGGATTCCTCTTAAAT 6684
923 TyrLeuIlePheAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsn 942
6685 AAATGGTGTGGAAAACTGGCTAGCCATATGCAGAAAGCTGAAAGCTGGATCCCTTCCTT 6744
943 LysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLysValAspProPheLeu 962
6745 ACACCTTATACAAAAGTTAACTCAAGTGAATTAAGACTTAAATATAAGACATAAAC 6804
963 ThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThr 982
6805 ATAAAAACCCCA-GAAGAAAACTAGGCAATACCATTCAGGATATGGCATGGCAAGAC 6863
983 IleLysThrLeuGluGluAsnLeuGlyIleThrIleGlnAspIleGlyValGlyLysasp 1002
6864 TTCATGACTAAACACCAAGCAATGGCAACAAAGCCAAATAGACAACTGGGATCTG 6923
1003 PheMetSerLysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeu 1022
6924 ATTAACCTATAGACTTCTGCACAGCAAAANAACTGTCATCAGAGTGAACAGCAAC 6983
1023 IleLysLeuLysSerPheCysThrAla-LysGluThrThrIleArgValAsnArgGlnPr 1042
6984 TACAGAAATGGAGAAAAATTTTGCATCTATCGATCTGACAAAGGCTTAATATCCAGAGAT 7043
1042 oThrTrpGluThrIlePheThrThrTyrSerSerAspLysGlyLeuIleSerArgIl 1062
7044 CTACGAGAAGCTTAAACAAATTTACAAGAAAAA ----AACACCCCTCAAAATATGGGC 7099
1062 eTyrAsnGluLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAl 1082
7100 AAAGGATATGAGCAGACACTCTCAAGAGAGACATTTATGCACCCCAACAAACATATCAA 7159
1082 aLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAlaLysLysHisMetLy 1102
7160 AAAAACCTCATCATCTATTGGTGGTAGAGAAATCAAAACAAACCAACACAGAGACATCCA 7219
1102 sLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrMetArgTyrHl 1122
7220 TCTCATCTAGTTAGATGTGATCACTAAAAAGTCAGGAAACAAACAAATGCTCGAGAGG 7279
1122 sLeuThrProValArgMetAlaIleIleLysLysSerGlyAsnAsnArgCysTrpArgGl 1142
7280 ATGTGGAGAAATAGGAACACTTTTCCACTGTTGTTGGGAATGTAAATTAGTTCAACCAT 7339
```

Db 1142 yCysGlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysLysLeuValGlnProLe 1162  
 QY 7340 GTGGAAGACAGTGTGGAGATTCCTTAAGATCTAGAACACAGAAATATCATTTGACCCAGC 7399  
 Db 1162 wTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAl 1182  
 QY 7400 ATCCCATCTAGTATATACCCAAAGAAATATAATCATCTCTATTATAAGACACATG 7459  
 Db 1182 aileProLeuLeuGlyIleTrpProAsnAspTrpLysSerCysCysTrpLysAspThrCy 1202  
 QY 7460 CACATATGTTTATTGAGCACTGATCATCATATAGCAAGACTTGGAAACCAACCAATG 7519  
 Db 1202 sThrArgMetPheIleAlaLeuPheThrIleAlaLysThrTrpAsnGlnProLysCy 1222  
 QY 7520 TCCATCAGTATGAGTGGATAAGAAACATGCGCATATATACCATGAATACTATGCG 7579  
 Db 1222 sproThrMetIleAspTrpIleLysLysMetTrpHisIleTrpMetGluTrpTyrAl 1242  
 QY 7580 ACCCATAAAG-GATGAGTTCATGCTCTTCCAGAGATATGATGAGCTGGAAACCAT 7638  
 Db 1242 aAlaIleLysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluThrI 1262  
 QY 7639 CATCTCAGCAACCTAAACACAGAAACAGAAACCAACACACCATGTTCTCCTACTGTAA 7698  
 Db 1262 eileLeuSerLysLeuSerGlnGluLysThrLysHisArgIlePheSerLeuIleGl 1282  
 QY 7699 TGGGAGT 7705  
 Db 1282 yGlyAsn 1284

## RESULT 7

ABGI4545  
 ID ABGI4545 standard; Protein; 1726 AA.

XX AC ABGI4545;  
 XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #14536.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.

XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC;  
 XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS78732.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX PS Claim 20; SEQ ID No 44904; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1726 AA;

## Alignment Scores:

Pred. No.: 0 Length: 1726  
 Score: 5086.00 Matches: 1154  
 Percent Similarity: 61.73% Conservative: 141  
 Best Local Similarity: 55.00% Mismatches: 298  
 Query Match: 9.72% Indels: 514  
 DB: 22 Gaps: 31

US-10-083-853B-2 (1-29921) x ABGI4545 (1-1726)

QY 1210 CATTCCTCCACTGAGTACCCAGTTCATCTC----- 1239  
 Db 8 HlsProProAlaGluTrpProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27  
 QY 1240 -----ACTGGGACTGGTTAGACATTGGGTGGCGAGCCAGGAGGTGACTGAA--- 1287  
 Db 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47  
 QY 1288 GCAGGTGGGTGGTCCCTCAGCCGCGAAGTCAAGGGGGTGGGGATCTCTCTCCGCCA 1347  
 Db 48 GlnGlyTrpAsnLysPro-----GlyIleGlnSerAspPro 59  
 QY 1348 GCCAAGGAGCCATGACAGACTGTACAGGAGGAGTAATGGTGCACTCTAGTCAGATAC 1407  
 Db 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69  
 QY 1408 CACTTTTCCCATAGTCTTTGCAACTGGCAGACAGGAGATTTCCCCAGTGCTATGCCA 1467  
 Db 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnHis 84  
 QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAACTGGCGGCCATT----- 1507  
 Db 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleIleGluGluPheIlePh 104  
 QY 1508 -----TGGACAGACACCGAGCTAGCCGCCACAGTTTATTTTTCATATCCCGCTGG 1557  
 Db 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerProLysSe 124  
 QY 1558 CGCTTGAATGCCAAGACAGAACCATTCACCTCCAGGGATCCAAAGTGGTCTGCTCAG 1617  
 Db 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136  
 QY 1618 T---GGGTCCACCCCATGGAGCCAGCTA---GCTAAGATCCAC----- 1657  
 Db 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysHis 156  
 QY 1658 -----TGGCTTGAATTTCTCTGCGCCAGCACAGCAGTC 1689  
 Db 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176  
 QY 1690 TGAGATTCA-----CCTGGG---ATGCTTGACTTGG 1718  
 Db 176 sGluValTyrSerMetArgThrMetGlyArgGlnProGlyThrAlaMetAspLeuAs 196

Qy 1719 TGAGGGAGGGGCTCTGCCATTCCTGAGGCTTCTAGTAGGGAGGCGGTTTACCCTCAA 1778  
Db 196 nAlaGlySerIleCysAsnVal----- 204  
Qy 1779 AGTGTAAACAAAGCTACTGGGAAGTTGAATGGGCGCCGCCACCGCAGCTCAGCAAGCGCG 1838  
Db 205 -----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr 222  
Qy 1839 CTGTGGCAAACTGCTCTAGATTCCTCTCTTTTGGGCGAGGTCTCTCTGAAAGAAAGG 1898  
Db 222 gValGly-----ACCTGGGG-----AlaTyrLeuAspIleArgG 231  
Qy 1899 CAGCAGCCCGCAGTCAGGACTTATAGATAAAACCCCATCTCCCTG---GGACAGAGA-- 1953  
Db 231 yTrpTyrIleLeuValLeuValAspLysGluLysLeuGlyIleProGlnLysG 251  
Qy 1954 -----ACCTGGGG-----AAGGGTGGCTGGGTGGCAGCT 1985  
Db 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270  
Qy 1986 TCTCAGACTTAAACATTCCTGCTGGAGGCTCTGAAGAGAGCAGCGGATCTCCAGCACA 2045  
Db 271 -----As 271  
Qy 2046 GCATTTGAGCTCTGATAGGAGCAGGCTGCTCTCAAGTGGGTCCCTGACCCCATGTA 2105  
Db 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283  
Qy 2106 TCCTGACTGGGAGACATCTCCATTAGGGGCCAATAGACATTTTCATACAGGAGACAGGT 2165  
Db 284 ----- 284  
Qy 2166 CTGGAGTGGACCTCCAGCAACTCCACGACCTGCAGCAGCGGCTGACTGTAGAA 2225  
Db 284 eTrpValGlyGln----- 288  
Qy 2226 GGAAGAGTAAACACAGAGGAATAGTATCAAC-----ATTAACAA 2267  
Db 289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305  
Qy 2268 AAGGACATCCACTCAGACGCCCTCTGAAGGTCAACAACATCAAGAGCAACAAAGTAAA 2327  
Db 305 gLysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValas 325  
Qy 2328 TAAACCAAAAGAGTGGGAAAAAC-CAGTGCAGAAACACTGAAATTTCCAAAACCCAGAA 2386  
Db 325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345  
Qy 2387 CTCTCTCTCAACCAAGGATCAACTCTCTCCAGCAAGGGAACAAACAGATGGA 2446  
Db 345 rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTrpThrG 365  
Qy 2447 GAATGAGTTTGAGGAATTGACAGAAAGTGGCTCAGAAAGTGGGTAATAACAACCTCCT 2506  
Db 365 uAsnAspPheAspLysLeuArgGluGluGlyPheArgArg-----SerAspTyrSe 382  
Qy 2507 CGAGCTAAAGAGAGCATCTTCAACCCCAATGCAAGAGAGCTTAAGAACCTTGAAAGAGGTT 2566  
Db 382 rGluLeuGlnGluGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe 402  
Qy 2567 AGATGAATTCCTACTAGATAATCATCTGATAGAGAGCAACATAATGACCTGATGGAGCT 2626  
Db 402 uAspGluTrpIleThrArgIleThrAsnAlaLysLysSerLysAspLeuMetGluLe 422  
Qy 2627 GAAAACCCAGACAAAGCTTCATGAAGCATACACAAAGTTCATAGCCAAATCGATCA 2686  
Db 422 uLysThrLysAlaArgGluLeuCysAspGluArgThrSerLeuSerSerGlnCysAsnG 442  
Qy 2687 AGCAGAAAGAGGATATCAGTGTAGATCAAAATTAATAAAGAAAGATGAGAAGACAA 2746  
Db 442 nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluGlu 462  
Qy 2747 GATTACAGAAAAAGAGTGAAGAGAACAAACAAAGCCTCCAGAAATTTAGGCACTATGT 2806

Db 462 sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnIleLysTrpAspTyrMe 482  
Qy 2807 GAAAAGACCAAAATCTACATTTGATTTGGTCCCCCAAGATGATGGGAGAGATGAACAA 2866  
Db 482 tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy 502  
Qy 2867 GTTGGAAAACACTCTTCCAGGTATTATCCAGGAGAATTTCCCAT-CTATCAGGCGCAGGC 2925  
Db 502 sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl 522  
Qy 2926 CAACATTCAATTCAGAAATATGGAGAACACACACATTAAGATCTCTCTCAGAGAAACAAT 2985  
Db 522 aAsnIleGlnIleGlnGluIleGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh 542  
Qy 2986 CCAAGACACATATCTTCAGATTCACCAAGGTGAAATGAAGGAAAAATTTGTTAAGGGC 3045  
Db 542 rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa 562  
Qy 3046 AGCCAGAGAAAGGTTGGGTTACCCACAAGGAGCAATCAGACTACAGCGGATCT 3105  
Db 562 lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe 582  
Qy 3106 CCCGGCAGAAACCTTACAAGCAGAGAGTGAGGCGCAATATTCACATTTCTTAAAGA 3165  
Db 582 uLeuAlaGluThrLeuGlnAlaArgArgGln----- 592  
Qy 3166 AAATAATTTTCAACCCCAAGATTTTCATATCCAGCCAAACCAAGCTTCTTAAGTGAAGAGA 3225  
Db 592 ----- 592  
Qy 3226 AATAAATCTCTACAGAGAAAGCAATGCTGCACAGATTTTGTCCACCAGCGCTCGCT 3285  
Db 592 ----- 592  
Qy 3286 TACAAGAGCTCCTGAAGGAAGCACCACACATGGAAGGAACAACCTGGTACCAGCCACTGCA 3345  
Db 592 ----- 592  
Qy 3346 AAAACATCCCAATTTGTAAGACCATTTGATGCTATGAAGAAGTGCATCAACTAACGGGC 3405  
Db 592 ----- 592  
Qy 3406 AAAATAACAGCTAGTGTCTCATATGCGAGGATCAATTCACACATAATATATTAACCTT 3465  
Db 592 ----- 592  
Qy 3466 AAATGTAAATGGCTAAATTTCCCAATTTAAAAGACACAGACTGGCAAAATTTGGATAAAGAG 3525  
Db 592 ----- 592  
Qy 3526 TCAAGACCCATCAGTGTGTATTTCAGGAGGCCCATCTCACATGAAGAGACACACATAG 3585  
Db 593 -----AspThrHisAr 596  
Qy 3586 GCTCAAAATAAAGGATGGAGGAAGATTTACCAGTAATATGGAACAAAAACAAAAAGC 3645  
Db 596 gLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAl 615  
Qy 3646 AGGGTTGCAATCTCTAGTCTCTGATAAAGCAGACTTTAAACCAACAAAGATCAAAAGAGA 3705  
Db 615 aGlyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs 635  
Qy 3706 CAAAGAGGCGCATTAATGCTAAAGGATCAATGGAAACAAAGAGCAAGTAACTATCTCT 3765  
Db 635 pLysGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluGluLeuThrIleLe 655  
Qy 3766 AATATATACATGACCCCAATACAGAGCACCAGGATTCATAAAGCAAGTCTTTAGAGACCT 3825  
Db 655 uAsnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe 675  
Qy 3826 ACAAGAGACTTTTACTCCACACAATAATAGTGGGAGTCTAAATAATAATATAGACACTT 3885  
Db 3885

Db	675	uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspH	689
QY	3886	TAACACCCACTGCCAATATTAGGCAGATCAATGAGACAGAAAATTAACAGGATATCCA	3945
Db	689	eAsnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG1	709
QY	3946	GGAGTTGAACAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCCACCC	4005
Db	709	nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyArgThrLeuHisPr	729
QY	4006	CAAAATCAACAGAAATACACTCTCTCAGCATCATCATACACCTATTTAAATGACCA	4065
Db	729	oLysSerThrGluAsnThrPhePheSerAlaProHisThrTySerLysIleAspHi	749
QY	4066	TGTAATTTAAGTAAACACTCCTCAGCAATGCAAAAGACAGAAATCCTAACAAACAG	4125
Db	749	sIleValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy	769
QY	4126	TCTCTCAGACTACAGTCAATCTATTTAGAACTCAGAATTAAAGAACTCAGTCAAAATCA	4185
Db	769	sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy	789
QY	4186	CACAACACTACATGGAACCTGAAACCTGCTCCTGAATCACTACTGGGTAATAACAAAT	4245
Db	789	sSerThrIleThrLysLeuAsnAsnLeuLeuAsnAspTyTrpValHisAsnGluH	809
QY	4246	GAAGCAAAATAAAGATGTTCTTTGAAACCAATGAGAAACACACAATGTACACGAA	4305
Db	809	tLysAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrTyrgLnaS	829
QY	4306	TCTCTGGGCGCATTTTAAAGCAGTGTGTAGAGGGAATTTATAGCACTAGATGCTACAA	4365
Db	829	nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisG1	849
QY	4366	GAGAAAGCAGCAATATCTAAATAGACACCTTACATCACAATTAAGAACTAGAGAA	4425
Db	849	yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLy	869
QY	4426	GAAGAGCAACAATTCAAAAGCTAGCAGAGACAGAAATAACTAAGATCAGAGCAGA	4485
Db	869	sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGlnIleThrLysIleArgAlaG1	889
QY	4486	ACTGAGAGATAGACACAAAAGCCCTTCAATAAATCAATGAATCCAGGAGCTGCT	4545
Db	889	uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrp	909
QY	4546	TTTTTGAAGATCAGCAAAAT---AGACCCTAGACAGACTAATAAGAAAGAAAGAG	4601
Db	909	hePheGluArgLysLeuAsnLysIleAspArgProLeuAlaArgLeuIleLysLysArgG	929
QY	4602	AGAAGAAATCAAGAGATGCAATAAAAAATGATAAGGGGATATCACCCAGATCCACAG	4661
Db	929	luLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspProThrG	949
QY	4662	AAATACAACTATATACAGAAATTTAATACACCTCTATGCAATAAATAGAAATC	4721
Db	949	luIleGlnThrThrIleArgGluTyTyLysHisLeuTyAlaAsnLysLeuGluAsnL	969
QY	4722	TAGAGAAATGGATAAATTCCTGGACACATATGTAGCCTGTATGGACCTTGGGGCAGA	4781
Db	969	eGluGluMetAspThrPheLeuAsp-----	977
QY	4782	ACAAAAGGGGTGAATGCAGAAATAAAGACAAAAGAGTATGTTTGAAGTAGG	4841
Db	977	-----	977
QY	4842	GGTCAAGGGGCACTTGCTCTAATGGACAAGGGCCCTGAGCTTTACACCCTCTGTA	4901
Db	977	-----	977
QY	4902	TTTATTAGGCAAAAGAGATACGAGAGGGTGCAGTTGGNAGAGAGCTGTAGTGC	4961
Db	977	-----	977
QY	4962	CAGAGTAGGCTGCAAGACTGCATTCTCAAAACAATAGGCTCTAGATGTCCAGTAGATA	5021
Db	977	-----	977
QY	5022	ACCTCAAGGAGCCAGTGCCAGGAGTGATGGCCCTCAGCAAACTTCTTAGGCGCAGACA	5081
Db	977	-----	977
QY	5082	GAAGTAAGTTTGCCCAACATTCGTATTCACGATAAACAGTTTGCTGTTGTATCAAGTAGC	5141
Db	977	-----	977
QY	5142	CTCCAGTGAATGTGAGTTGGTCATGATCCCTTTGGCCTTTTGGCTCCCAAAACACAT	5201
Db	978	-----	978
QY	5202	ACACCTCTCAAGACTAAACAGGAAGAGTCAAAATCCTCTGAATATACCAATACCAAGTT	5261
Db	979	yrThrLeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGlys	999
QY	5262	CTAAATTTGAAGCAGTAATTCATAGCTACCAACCAAAAAGTCCAGGACGACGAT	5321
Db	999	erGluIleValAlaIleIleAsnSerLeuProThrLysLysSerProGlyProaspGlyP	1019
QY	5322	TCAGAGCAAAATTTACAGAGGTACAAAGAGAGCTGGTACTATTCTCTGAAACTAT	5381
Db	1019	heThrAlaGluPheTyrgLysGluGluLeuValProPheLeuLeuLysLeuP	1039
QY	5382	TCCAAAAAATAGAA---AATGGGAATCCTCCTCACTCACTTTTACGAGCGCAGCATATC	5438
Db	1039	heGlnSerIleGluLysGlu-GlyIleLeuProAsnSerPheTyrgLysThrSerIleIle	1058
QY	5439	CTGATACCAAAACCTAGCAGTGACACACAAAAGAGAGAAATTTACAGCCCATATCCCTG	5498
Db	1059	LeuIleProLysProGlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeu	1078
QY	5499	ATGAACATTTGATGTGAAAATCCTCAATAAAATCTGCGAAACCAACCAATCCAGCAGCATC	5558
Db	1079	MetAsnIleAspAlaLysIleLeuAsnLysIleLeuAlaAsnArgIleGlnHisIle	1098
QY	5559	AAAAAGCTTATCTACATGATCAGTTGGCGTCATCCCTGGATGTCAGAGCTGGTTCAA	5618
Db	1099	LysLysLeuIleHisHisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsn	1118
QY	5619	ATATGCAATCAATAAATGTAGGCCATCACATAAACCAACCAATGACAAAAACACATG	5678
Db	1119	IleArgLysSerIleAsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMet	1138
QY	5679	ATTATCTCAATPAGATGCAAAAAGCCCTTCTCAAAATTCACAGCCCTTCATGCTAAAA	5738
Db	1139	IleIleSerIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLys	1158
QY	5739	ATTCCTCAGTAACCTAGGTATCGATGGAATGTATCTCAAAATTAAGAGCTATTATATAC-	5797
Db	1159	ThrLeuAsnLysLeuGlyIleAspGlyThrTyPheLysIleIleArgAlaIleTyArgP	1178
QY	5798	AAACCCAGCAGCAATATCATACTGAATGGGCAAAACCTGGAAGCATTCCTCTTGAGAAT	5857
Db	1179	LysProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThr	1198
QY	5858	GGCAACAGACAGGATGCCCTCTCTCAGCATCTCTATTCAAGATACTATTGAGAGTCTG	5917
Db	1199	GlyThrArgGlnGlyCysProLeuSerProLeuLeuPheAsnIleValLeuGluValLeu	1218
QY	5918	GCCAGGGCAATCAGCAATAGAAATAAAGGTATTCATAATAGAAAGAGAGAGATC	5977
Db	1219	AlaArgAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluGluVal	1238
QY	5978	ATATTGCTCTGTTTGGCAGATCATCTTTTATATATTTAGAAAACCCCTCTCTCAGC	6037
Db	1239	LysLeuSerLeuPheAlaAspMetIleValTyLeuGluAsnProIleValSerAla	1258







Alignment Scores:

Pred. No.: 0 Length: 2030  
Score: 5067.50 Matches: 1100  
Percent Similarity: 65.15% Conservative: 128  
Best Local Similarity: 58.36% Mismatches: 253  
Query Match: 9.68% Indels: 412  
DB: 22 Gaps: 14

US-10-083-853B-2 (1-29921) x ABG24249 (1-2030)

QY	2188	TCACGACGACCTGCAGCAGAGCGGCTGACTCTTAGAGGAAAGTAACAACAAAGG	2247
DB	446		463
QY	2248	AATAGTATCAACATTAACAAAAGGACATCCACTCAGAGACCCCTCTGAAGGTCAACAA	2307
DB	464		477
QY	2308	CATCAAGACCAAGGTAAATAAACCAGATGGGAAAC-CAGTGCAGAAACACT	2366
DB	478		497
QY	2367	GAAATTCAAAACCAAGACCTCTCTCAACCAAGGATCAACAACCTCCGCCAGCA	2426
DB	498		517
QY	2427	AGGGAACAAACAGATGAGATGAGTTGAGGAATTCACAGAGTAGGCTTCAGAAGG	2486
DB	518		537
QY	2487	TGGGTAAATCAACACTCTCCGAGCTAAAGGAGCATGTTCTAACCCAAATGCAAGGAAGCT	2546
DB	538		554
QY	2547	AGAACTTGAAGAGGTAGATGAATTCCTAACTAGATATATCAGTGTAGAGAGAAC	2606
DB	555		574
QY	2607	ATAATGACCTGATGAGCTGAAAACGCAAGACAGAACTTCATGAACATACACAAGC	2666
DB	575		594
QY	2667	TTCATAGCAATCGATCAACGAGAGAAAGATATCAGTGTGAAGATCAATTAAT	2726
DB	595		614
QY	2727	AAAAGAACTGAGACAGCAAGATTACAGAAAAGAGTCAAGAAAACCAACAAAGCCCTC	2786
DB	615		634
QY	2787	CAAGATATGGGACTATGTGAAGACCAATCTACATTTGATGGTCTCCCAAGT	2846
DB	635		654
QY	2847	GATGGGAATGGAATCAAGTTGGAACACACTCTCAGGGATTTATCCAGAGAAATTC	2906
DB	655		674
QY	2907	CCCAT-CTATCAGGCGAGCCCAACATTCAAATTCAGGAATATGGAGACACCAATAAGA	2965
DB	675		694
QY	2966	TACTCTCAGAGAACCAATCCAGACACATATCTTCAGATTCACCAAGTTGAAATG	3025
DB	695		714
QY	3026	AAGGAAAAATGTTAAGGCGACGCCAGAGAGAAGTTGGGTTTACCCCAAGGAGGCCA	3085
DB	715		725
QY	3086	ATCAGACTAACAGCGGATCTCCGGCAGAAACCCCTACAGCCAGAGAGTGAGGGCCA	3145
DB	725		745

QY	3146	ATATCCACATCTTTAAAGAAAATAATTTTCAACCCAGAAATTTTCATATCCAGCCAAACCA	3205
DB	725		745
QY	3206	AGCTTCTTAAGTGAAGGAGAAATAAATCTCTTACAGAGAAGCAATGCTGACAGATTTT	3265
DB	725		745
QY	3266	TGTCACCACCGGCTGCTTACAGAGCTCTTGAAGAGCAACCAACATGGAAGGAAC	3325
DB	725		745
QY	3326	AACTGGTACCAGCCACTGCAAAAACATCCAAAATTTGCTTAAAGACCATTTGCTATGAAGA	3385
DB	725		745
QY	3386	AAGTCATCAACTAAGGGCAAAATAACAGCAGTAGTGTCTATATGCGCAGATCAAAATTC	3445
DB	725		745
QY	3446	CACATAATAATTAACCTTAAATGTAAATGGCTAAATTCCTCAATTTAAAGACACAGA	3505
DB	725		745
QY	3506	CTGGCAATTTGGATAAAGAGTCAAGACCCATCAGTGTGTATTTCAGGAGGCCCATCTC	3565
DB	725		745
QY	3566	ACATGAAAAGACACACATAGGCTCAAAATAAAGGATGGAGGAAGATTTACCAAGTAAT	3625
DB	725		745
QY	3626	GGAAACAAAAAAGCAGGGTTGCAATCTCTGATCTGATAAAGACAGACTTTAA	3685
DB	726		745
QY	3686	CCAACAAAGATCAAAAGAGACAAAGAGCCATTACATAATGTAAGGATCAATGGA	3745
DB	729	ProThrThrIleLysArgAspLysGluGlyHisTyrIleMetValLysGlySerIleGln	748
QY	3746	CAAGAAGACTTAATCTCTTAATATACATGACCAATACAGGACCCAGATTCATA	3805
DB	749	GlnGluLeuThrIleLeuAsnIleTyrAlaProAsnThrGlyAlaProArgPheIle	768
QY	3806	AAGCAAGTTCTTAGAGACCTTACAAAGAGACTTTGACTCCACACAATAATAGTGGAGTC	3865
DB	769	LysGlnValLeuSerAspLeuGlnArgAspLeuSerHisThrLeuIleMetGly	787
QY	3866	TAAATAATAATAGACACTTTAACCCCTGCGCAATATTTAGGAGATCAATGAGACAG	3925
DB	788		802
QY	3926	AAATTAACAAGGATATCCAGGAGTTGAAGTCTGAGCTCGACCAAGCGACCTAATAGAT	3985
DB	803	LysValAsnLysAspThrGlnGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAsp	822
QY	3986	ATCTACAGAACTCCCAACCCCAAAATCAACAGAAATATACACTCTTCTCAGCATCACATTAC	4045
DB	823	IleTyrArgThrLeuHisProLysSerThrGluTyrThrPhePheSerAlaProHisHis	842
QY	4046	ACCTATTTTAAATTTGACCATGTAAATTTAAGTAAACACTCTCTCAGCAAAATGCAAAAGA	4105
DB	843	ThrTyrSerLysIleAspHisIleLeuGlySerLysAlaLeuLeuSerLysCysLysArg	862
QY	4106	ACAGAAATCTTAACAAACAGTCTCTCAGACTACAGTGCATCTATTATTAGAACTCAGAAAT	4165
DB	863	ThrGluIleThrAsnTyrLeuSerAspHisSerAlaIleLysProGluLeuArgIle	882
QY	4166	AAGAACTCACATAAAATCACAACTACATGAACTGAACTGAACTGCTCTGATGATGAC	4225
DB	883	LysLysLeuThrGlnAsnCysSerThrThrTrpLysLeuAsnLeuLeuLeuAsnAsp	902

QY 4226 TACTGGGTAATAACAAATGAAGCAAAATAAGATGTTCTTTTGAACCAATGAGAAC 4285  
Db TyrTrpValHisAsnGluMetLysAlaGluIleLysMetPheGluThrAsnGluAsn 922  
QY 4286 AAGACACAAATGTACCAGAACTCTGGGGCATATTTAAAGCAGTGTGTAGAGGAATTT 4345  
Db LysAspThrThrTyrglnAsnLeuTrpAspThrPheLysAlaValCysArgGlyLysPhe 942  
QY 4346 ATAGCACTAGATGCTCAACAGAGAAAGCAGCAATATCTAAATAGACACCTTACATCA 4405  
Db IleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIleAspThrLeuThrSer 962  
QY 4406 CAATTAAAGAACTAGAGAAAGAGCAAAATAATTCAAAGCTAGCAGAGACAGAA 4465  
Db GlnLeuLysGluLeuLysGlnGluGlnThrHisSerLysAlaSerArgArgGlnGlu 982  
QY 4466 ATACTAGATCAGACGAGAACTGAAGGAGATAGACAGACAAAAGCCCTTCAATAAAT 4525  
Db IleThrLysIleArgAlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLys-I 1002  
QY 4526 CAATGAATCCAGGAGCTGGTTTTTGAAGAGATCAGCAAAAT---AGACCACCTAGACAG 4581  
Db eAsnAspSerArgSerTrpPheGluLysIleAsnLysIleAspArgProLeuAlaAr 1022  
QY 4582 ACTAATAAAGAAAGAGAGAGAAATCAAGAGATGCAATAAAATAATGATAAGGGGA 4641  
Db gProLysLysLysLysArgGluLysAsnGlnIleAspAlaIleLysIleAspLysGlyAs 1042  
QY 4642 TATCACCACCATCCACAGAAATACAACTATTATCAGAGAAATATATAACACCTCTA 4701  
Db pIleThrThrAsnProThrGluIleGlnThrIleArgGluTyrrLysArgLeuTy 1062  
QY 4702 TGCAAATAAAGTACAGAAATCTAGAGAAATGGATAAAATTCCTGGACACATATGTAGCCTG 4761  
Db rAlaAsnLysLeuAlaAsnLeuGluMetAspLysPheAsp 1077  
QY 4762 TATGGACCTTGGGGACAGAACAAAGGGGTGAATGCAGAAATAAAGACAAAGACAAA 4821  
Db 1077 ----- 1077  
QY 4822 AGAGTATGTTTGAAGTAGGGTCAGGGGGCAACTTGCTCTAATGGACAGGCCCTGA 4881  
Db 1077 ----- 1077  
QY 4882 GCTTTACACCCTCTGTATTATTAGCAAAAGATAGCAGAGGGGTGAGTTGGAAG 4941  
Db 1077 ----- 1077  
QY 4942 AAGAGGTGAGTGTAGGTCAGAGTAGGCCCTGCAAGACTGCATTCCTCAAAACAATAGGC 5001  
Db 1077 ----- 1077  
QY 5002 TCTAGATGTCCTCAGTAGTACCTCAAGGAGCCAGTGCCAGGAGTGATGGCCCTCAGCA 5061  
Db 1077 ----- 1077  
QY 5062 AACCTTCTAGGCGACGACAGAGTAAGTTTGCCACATCTGTATTACAGATAAAGCT 5121  
Db 1077 ----- 1077  
QY 5122 TTGCTGTTGATCAAGTAGCCTCAGTGGAATGCTGAGTGGTCATGATCCCTTTGGCCT 5181  
Db 1077 ----- 1077  
QY 5182 TTTTGGCTCCCAAAACACATACCCCTCTCAAGACTAAACAGGAGAGTCAAAATCCCT 5241  
Db 1078 -----ThyrrThrLeuProArgLeuAsnGlnGluValGluSerLe 1092  
QY 5242 GAATATACCGATACAAAGTTCTAAATTAAGCAGTAAATGATAGCCTTACCAACCAAAA 5301  
Db uAsnArgProLysThrGlyAlaGluIleValAlaIleIleAsnSerValProThrLysLy 1092  
QY 5302 AAGTCCAGGACCGAGGATTCACAGCCAAATTCACAGAGGTACAAAGAGAGCTGGT 5361

Db sSerProGlyProAspGlyPheThrAlaLysPheTyrglnSerTyrrLysGluGluLeuVa 1132  
QY 5362 ACTATTCTCTCTGAACATATTCCAAAATAATAGAA---AATGGGAATCCTCCCTAACTCAT 5418  
Db 1132 1ProPheLeuLeuLysLeuPheGlnSerIleGluLysGlu-ArgIleLeuProAsnSerP 1152  
QY 5419 TTTACGAGGCGCAGCATCTCTGATACCAAAACCTAGCAGTGACACACAAAGAGAA 5478  
Db 1152 heTyrgluAspSerIleIleLeuIleProLysProGlyArgAspThrThrLysLysGluA 1172  
QY 5479 ATTTACGCGCCATATCCTGATGAACATTTGTAATAATCTCAATAAATACTGGCAA 5538  
Db 1172 snPheArgProIleSerLeuMetAsnIleAspAlaLysIleLeuAsnLysIleLeuAla 1192  
QY 5539 ACCAAATCCAGCAGCACATCAAAAGCTTATCTACCATGATGATGGCGTCATCCCTG 5598  
Db 1192 snGlnIleGlnGlnHisIleLysLysLeuIleHisAspGlnValGlyPheIleProG 1212  
QY 5599 GGATGCAAGGCTGCTTCAAAATATGCAAAATCAATAAATGATAGGCATCACATAACAGAA 5658  
Db 1212 lYmetGlnGlyTrpSerThrIleHisLysSerIleAsnValIleGlnHisIleAsnArgT 1232  
QY 5659 CCAATGACAAACACATGATTATCTCAATAGATGAGAAAGCGCTTGTCTCAAAATTC 5718  
Db 1232 hrLysAspLysLysHisMetIleIleSerIleAspAlaGluLysAlaPheAspLysIleG 1252  
QY 5719 AACAGCCTTCATCTCAAAATCTCAGTAACTAGGTATGATGATGATGATGATGATGATG 5778  
Db 1252 lGlnProPheMetLeuLysThrLeuAsnLysLeuGlyIleAspGlyThrTyrrLysLysI 1272  
QY 5779 TAATAAGAGCTATTATATAC-AAACCCACAGCAATATCATACTGAATGGGCAAAACTGG 5837  
Db 1272 leIleSerAlaIleTyrrAspLysProThrAlaAsnIleIleLeuAsnGlyGlnLysLeuG 1292  
QY 5838 AAGCATTCCTTTGAGAACTGGCACAAGACAGATGCCCCCTCTCTCACCACCTCTATTCA 5897  
Db 1292 luAlaLeuProLeuLysThrGlyThrArgGlnGlyCysProLeuSerProLeuLeuPheA 1312  
QY 5898 AGATCTATTGGAAGTTCTGGCCAGGCAATCAGGCAATAGAAAGAAATAAGGATATTC 5957  
Db 1312 snIleValLeuGluValLeuAlaArgAlaIleArgGlnGluLysGluIleLysGlyIleG 1332  
QY 5958 AAATAGAAAGAGAGAAAGTATATGTTCTGTTTGCAGATGACATGTTGTATATATTAG 6017  
Db 1332 lNLeuGlyLysGluGluValLysLeuSerLeuPheAlaAspMetIleValTyrrLeuG 1352  
QY 6018 AAAACCCCATGCTCTCAGGCCCAAAACTCCTTAAGCTGATAGCAACTTCAGCAAGCTCT 6077  
Db 1352 luAsnProIleValSerValGlnAsnLeuHisLysLeuIleSerAsnPheSerLysValS 1372  
QY 6078 CAGGACACAAATCAATGTGCAAAATCACAAGCATTTTATACGCCAATAAATAGACAAA 6137  
Db 1372 erGlyTyrrLysIleAsnValGlnLysSerGlnAlaPheLeuTyrrThrAsnAsnArgGluT 1392  
QY 6138 CAGAGACCAATCATGAGTCACTCTCATTCACAATTTGCTACAAAGAGATAAAATACC 6197  
Db 1392 hrGluAsnGlnIleMetSerGluLeuProPheThrIleAlaSerLysArgIleLysTyrrL 1412  
QY 6198 TAGGAATACAACTTACAGGACAGCTAGGAACCTTCAAGGAGAACTACAAACCACTGA 6257  
Db 1412 euGlyIleHisLeuThrArgAspValLysAspLeuPheLysGluAsnTyrrLysProLeuL 1432  
QY 6258 TCAAGGAAATAAGAGAGACACAAACAAATGGAATAATCCATGCTCACAGATGATAA 6317  
Db 1432 euAsnGluIleLysGluAspThrAsnLysTyrrLysAsnIleProCysSerTrpValGlyA 1452  
QY 6318 GAATCAT-----GAAATGCCATCTGCCCAAGTAATATATAGATTTCAGTCTCCTACC 6370  
Db 1452 rgIleAsnIleValLysMetAlaIleLeuProLysValIleTyrrArgPheAsnAlaIleP 1472  
QY 6371 CCATCAGCTACCATTCACCTTCTTTCACAGATTGGAAACAACTTAAATTTTCATAT 6430

```
Db 1472 roileuLysLeuProMetThrPhePheThrGluLeuGluLysThrThrLeuLysPheIleT 1492
QY 6431 GGAACCAAAAAAGAGCCACAGACCAAGACAACTTTAAGCAAAAAAGCAAAAGCTGGA 6490
Db 1492 rPAsnGln- LysArgAlaHisIleAlaLysSerIleLeuSerGlnLysAsnLysAlaGly 1511
QY 6491 GGTATCATGCTACCTGACTTAAACTATATACTATAGGCTACAGTACCAAAAACACTGCATGG 6550
Db 1512 GlyIleMetLeuProAspPheLysLeuThrHisLysAlaThrValThrLysThrAlaTrp 1531
QY 6551 TACTGGTACCAAAACAGATATATACCAACATCGACAGACAGACACCTCAAAAATTACA 6610
Db 1532 TyrTrpTyrGlnAsnArgAspIleLeuGlnTrpAsnLysThrGluProSerGluIleThr 1551
QY 6611 CT-GCAATCATCATCATCTGATCTTGCACAACTGCACAAAACAAAGCAATGGAAAAG 6669
Db 1552 ProHisIleTyrAsnTyrLeuIlePheAspLysProGluAsnLysGlnTrpGlyLys 1571
QY 6670 GATTCCTTATTAATAATGGTGGTGGAAAACTGGCTAGCCATATGCAGAAAGCTGAAA 6729
Db 1572 AspSerLeuPheAsnLysTrpCysTrpGluAsnTrpLeuAlaIleCysArgLysLeuLys 1591
QY 6730 CTGGATCCCTCTCTACACCTTATACAAAGTTAACTCAAGATGAATTAACACCTTAAT 6789
Db 1592 LeuAspProPheLeuThrProTyrThrLysIleAsnSerArgTrpIleLysAspLeuAsn 1611
QY 6790 ATAAGACATAAAACCAACCAACCA- GAAGAAAACTAGGCAATACCATTCAGGATATG 6848
Db 1612 ValArgProLysThrIleLysThrLeuGluLysLeuGlyIleThrIleGlnAspIle 1631
QY 6849 GACATGGGCAAGACTTCATGACTAAACACCAAGCAATGGCAACAAAGCCCAAAATA 6908
Db 1632 GlyIleGlyLysAspPheMetSerLysThrProLysAlaMetAlaThrLysAspLysIle 1651
QY 6909 GACAAGTGGGCTCATTAACATATAGAGCTTCTGCACAGCAAAAACAACTGTCATCAG 6968
Db 1652 AspLysTrpAspLeuIleLysLysSerPheCysThrAla- LysGluThrIleAla 1671
QY 6969 AGTGAACAGCAACCTCAGAAATGGGAGAAAATTTTGCATCTATCGATCTGACAAAGG 7028
Db 1671 gValAsnArgGlnProThrLysTrpGluGluIlePheAlaThrTyrSerSerAspLysG 1691
QY 7029 CTAATATCCAGAGTCTAGGAAGAATTAAACAATTTACAAGAAAA- -AACAAACC 7084
Db 1691 yLeuIleSerArgIleTyrAsnGlyLeuLysGlnIleTyrLysLysThrAsnAsnPr 1711
QY 7085 CGTCAAAATATGGCAAGGATATGAGCAGACACTTCTCAAAAGAGACATTTATGCACG 7144
Db 1711 oileLysTrpTrpAlaLysAspMetAsnArgHisPheSerLysGluAspIleTyrAlaAl 1731
QY 7145 CAACAAACATATGAAAAAACCTCATCATCTGCTGTAGAGAAATGCAAAAACAAAC 7204
Db 1731 aLysLysHisValLysLysCysSerSerSerLeuAlaIleArgGluMetGlnIleLysTh 1751
QY 7205 CACAGTGACATACCATCTCATGCTAGTATAGATGGTGATCACTAAAGATCAGGAACAA 7264
Db 1751 rThrMetArgTyrHisLeuThrProValArgMetAlaIleIleLysLysSerGlySerAs 1771
QY 7265 CAAATCGCTGGAGAGATGGGAGAAATAGGACACATTTCCACTGTGGTGGGATGTAA 7324
Db 1771 nArgCysTrpArgGlyCysGlyGluIleGlyThrLeuLeuHisCysTrpTrpAspCysL 1791
QY 7325 ATTAGTTCAACCATTTGTGAAGACAGCTGGGAGATTCCTTAAGGATCTAGAACCAAGAT 7384
Db 1791 sleuValGlnProLeuTrpLysSerValTrpArgPheLeuArgAspLeuGluLeuGluI 1811
QY 7385 ATCATTTGACCAAGCAATCCCATTTACTGAGTATATACCAAGAGAAATATAATTCATCT 7444
Db 1811 ePropPheAspProAlaIleProLeuLeuGlyIleTyrProAsnAspTyrLysSerCys 1831
QY 7445 TTATAAGACACATGCACACATATGTTT-----AT 7474
Db 1831 sTyrLysAspThrCysThrLeuSerLeuProGluIleArgAsnCysProThrLysSerVa 1851
```

```
QY 7475 TGCAGCACTGATCAACAATAGCAAGACTTGGAAACCAACCAAAATGCTCCATCATGATAGA 7534
Db 1851 lProValProPheSerValAspGluThrGlnTyrHisValGluAspThrSerProValse 1871
QY 7535 CTGGATAAAGAAAAATGCG- - - - -CACATATACACCATGGAATACTATGACAGCCATA- - 7586
Db 1871 rTrpGlnValLysValLeuGlyLysHisIleHisGlnGlySerSerTyrGluAspValTr 1891
QY 7587 - - - - -AAAAGGATGAGTTTCATGCTCTTGCAGAGATATGATGAAGCTGGAAACCATC 7639
Db 1891 pLeuProGlnArgMetGlnGly- LysLeuThrAsnArgLysAsp- - - - -IleH 1907
QY 7640 ATTCTCAGCAAACTAACACACAGAAACCAACCAACCAACCATGTT- - - - -CTCACATT 7693
Db 1907 lThrLysAsnProSerValHisHisHisGlnArgProLysValAspLysThrThrL 1927
QY 7694 GTAAGTGGGAGTTGAACATGAGACACATGACACACAGGAGGAGGACACATCACACACAG 7753
Db 1927 yMetGlyLysLysGlnAsnArgLysThrGlyAsnSerLysLysGlnSerAlaSerProp 1947
QY 7754 GTCCT 7758
Db 1947 roPro 1948
RESULT 9
ABG10129
ID ABG10129 standard; Protein; 2764 AA.
XX
AC ABG10129;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10120.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS74316.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 40488; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
```



```
Db 415 rAlaIleLysLeuGluLeuArgIleLysAsnLeuAsnGlnSerArgSerThrThrTrpLys 435
Qy 4201 ACTGAACAACCTCTCTGAATGACTACTGGGTAAATAACAAATGAAGGCAAAATAAA 4260
Db 435 sLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGlnuMetLysAlaGluIleLys 435
Qy 4261 GATGTTCTTTGAAACCAATGAGAACAAAGACACAATGTACGAGAATCTCTGGGCGATATT 4320
Db 455 sMetPhePheGluThrAsnGluAsnLysAspThrThrTyrGlnAsnLeuTrpAspAlaPh 475
Qy 4321 TAAAGCAGTGTGTAGAGGGAATTTATAGCAGTATAGTGCCTACAGAGAAAGCAGGAAT 4380
Db 475 eLysAlaValLysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluAr 495
Qy 4381 ATCTAAATAGACACCTTAACATCACAAATTAAGAAGAACTAGAGAAGAAAGACAAACAA 4440
Db 495 gSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlyGlnGluGlnThrHI 515
Qy 4441 TTCAAAAGCTPAGCAGAGCAAGAAATAACTAAGATCAGACGAGAATCTGAAGGAGATAGA 4500
Db 515 sSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGI 535
Qy 4501 GACAAAAAGCCCTTCAATAATCAATCAATCAATCCAGGAGCTGGTTTTTGAAGAATCA 4360
Db 535 uThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePheGluArgIleA 555
Qy 4561 GCAAAAT-----AGACCACCTAGACAGACTAATAAGAAGAAAGAGAGAAATCAAGAG 4616
Db 555 snLysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluGluAsnGlnIleA 575
Qy 4617 ATGCAATAAAAAATGATAAGGGGATATCACCCGATCCACGAAATCAACAACATTA 4676
Db 575 spThrIleLysAsnAspLysGlyAspIleThrAspProThrGluIleGlnThrThrI 595
Qy 4677 TCAGAGATATTATAACACCTCTATGCAATAATACTAGAAAATCTAGAGAAATGGATA 4736
Db 595 leArgGluTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGluGluMetAspL 615
Qy 4737 AATTCCTGGACACATATGTAGCCTGTGTAGCACCTTGGGACCTGGGGGACAGAACAAAAGGGGTGAA 4796
Db 615 yspHeLeuasp-----Thr 618
Qy 4797 TGCAGAAATAAAAGACAAAGACAAAGAGATATGTTTGAAGTAGGGGTACAGGGGCAACT 4856
Db 618 ----- 618
Qy 4857 TGCCCTCTAATGGACAGGGCCCTGAGCTTTACACCACCTCTGTATTATTAGGCAAAAG 4916
Db 618 ----- 618
Qy 4917 AGATAGCGAGAGGGTCAGTTGGAAGAGAGGTGAGTGTGTAGGTCCAGAGTAGGCCGTGCA 4976
Db 618 ----- 618
Qy 4977 AGACTGCATTCCTCAAAACATAGGCTCTAGATGTCCAGTAGATACCTCAAGGAGCCAG 5036
Db 618 ----- 618
Qy 5037 TGCCAGGAGTGTAGTGGCCCTCAGCAAAACCTTCTAGGCGAGGCACAGAACTAAGTTTGGCC 5096
Db 618 ----- 618
Qy 5097 ACATTCCTGTATTCACGATAAAGTGTGCTGTTGTTGATCAAGTAGCCTCCAGTGAATGCT 5156
Db 618 ----- 618
Qy 5157 GAGTTGGTCATGATFCCCTTTGGCCCTTTTGGCTFCCCAAAACATACACCTCTCAAGAC 5216
Db 619 -----ThrTyrThrLeuProArgL 625
Qy 5217 TAAACAGAGAGAGTCAATCCCTGAATATACAGTAAACAAGTTCTAAATTTAGAGCAG 5276
Db 625 euAsnGlnGluValGluSerLeuAsnArgProIleThrGlyAlaGluIleValAlaI 645
```

```
5277 TAATTGATAGCTTACCAACCAAAAAAGTCCAGGACGAGGATTCACAGCCAAATTC 5336
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 leIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheThrAlaGluPhe 665
Qy 5337 ACCAGAGGTACAAGAGAGCTGGTACTATTCTCTCGAAACATATTCACAAAATAGAA - 5395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
665 yrglnArgTyrLysGluGluValProPheLeuLeuLysLeuPheGlnSerIleGlu 685
Qy 5396 --AATGGGATCCCTCACTACTTACAGAGCCGAGCATCATCTGATACCAAAACCT 5453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
685 ysglu-GlyIleLeuProAsnSerPheTyrGluAlaSerIleIleLeuIleProLysPro 704
Qy 5454 AGCAGTGACACACAAAGAGGAAATTTCCAGGCCATATCCCTGATGAACATTGTGTG 5513
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 GlyArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetAsnIleAspAla 724
Qy 5514 AAAATCCTCAATAAATCTGGCAACCAATCCAGCAGCACATCAAAAAGCTTATCTAC 5573
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 LysIleLeuAsnLysIleLeuAlaAsnArgIleGlnGlnHisIleLysLysLeuIleHis 744
Qy 5574 CATGATCAAGTTCGCTCATCCCTGGGATCGAAGCTGGTCAAAATATCAATCAATA 5633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 HisAspGlnValGlyPheIleProGlyMetGlnGlyTrpPheAsnIleArgLysSerIle 764
Qy 5634 AATGTAGGCCATCACATAAAGACCAATGACAAAAACCAATGATTTATCTCAATAGAT 5693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
765 AsnValIleGlnHisIleAsnArgAlaLysAspLysAsnHisMetIleIleSerIleAsp 784
Qy 5694 GCAGAAAAGCCCTTTGCAAAATTCACAGCCCTTCATGTAATAAATTCAGTAACATA 5753
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
785 AlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLeuAsnLysLeu 804
Qy 5754 GSTATCGATGAATGTATCTCAAAATAAAGAGCTATTATTATAC-RAACCCACAGCCAAT 5812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 GlyIleAspGlyThrTyrPheLysIleIleArgAlaIleTyrAspLysProThrAlaAsn 824
Qy 5813 ATCATACTGAATGGGCAAAACCTGGAGCATTCCTTTGAGAACTGGCACAAGACAGGA 5872
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
825 IleIleLeuAsnGlyLysLysLeuGluAlaPheProLeuLysThrGlyThrArgGlnGly 844
Qy 5873 TGCCCTCTCTCACCACCTCTATTTCAGATACTATTGGAGTTCTGGCCAGGCAATCAGG 5932
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
845 CysProLeuSerLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArg 864
Qy 5933 CAATAGAAAGAAATAAAGGTATTCAATAGAAAGAGAGAGAACTCATATTGTCTCTTT 5992
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 GlnGluLysGluIleLysGlyIleGlnLeuGlyLysGluValLysLeuSerLeuPhe 884
Qy 5993 GCAGATGACATGTTGTATATTTAGAAAACCCCATCGCTCAGCCGCAAAACCTCCTTAAG 6052
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
885 AlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnAsnLeuLys 904
Qy 6053 CTGATAGCAACTTCAGCAAAAGTCTCAGGACACAAATCAATGTGCAAAAATCACAGCA 6112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 LeuIleSerAsnPheSerLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAla 924
Qy 6113 TTCCTTATACCCCAATATACACAAACAGAGAGCAAAATCATAGTGAACCTCATTCACA 6172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
925 PheLeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThr 944
Qy 6173 ATTGCTACAAGAGAATAAATACCTAGGAATACACTTACAAAGGACAGCTGGAACCTC 6232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
945 IleAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspLeu 964
Qy 6233 TTCAGGAGAACTACAAACCTGATCAAGGAATATAGAGAGACACAAACAATGGAAA 6292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
965 PheLysGluAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTrpLys 984
Qy 6293 AACATTCCTGCTCACAGATAGTAAGATCAT-----GAAATGCCATCTGCCCAA 6345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
985 AsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaIleLeuProLys 1004
```

QY	6346	CTAAATTATAGATTACGTGCTACCCCTCCATCAAGCTACCATGACTTCTTCCACAGAAATTG	6405
Db	1005	VallileTyArgPheAsnAlaileProileLysLeuProMetThrPhePheThrGluLeu	1024
QY	6406	GAARAAACACTTAAATTTTCATATGGAACCAAAAGAGCCACAGACCAAGACAAAT	6465
Db	1025	GluLysThrThrLeuLysPheileTrpAsnGln-LysArgAlaArgileAlaLysSeril	1044
QY	6466	CTTAAGCAAAAAGCAAAAGCTGGAGGTATCATGCTACTGACTTAAACTATATATAA	6525
Db	1044	eLeuSerGlnLysAsnLysAlaGlyLleThrLeuProAspPheLysLeuTyTyLys	1064
QY	6526	GGCTACAGTAACCAAACTGCATGCTACTGTGTACAAACAGATATATAGACCAATGAA	6585
Db	1064	sAlaThrValThrLysThrAlaTrpTyTrpThrGlnAsnArgAspLysGlnTrpAs	1084
QY	6586	CAGAACAGAGACCTCAGAAATACACT-GCAATCTACATCATCTGATCTTTGCAACAC	6644
Db	1084	nArgileGluProSerGluileThrProHisThrTyTrpAsnTyTrpLeuilePheAspLysPr	1104
QY	6645	TGACAAAACCAAGCAATGGAAGGATTCCTCTATTTAATAATGCTGTGGAACAACTG	6704
Db	1104	oGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTr	1124
QY	6705	GCTAGCCATATGCAAGAAAGCTGAACCTGGATCCCTTACACCTTATACAAAGTTAA	6764
Db	1124	pLeuAlaileCysArgLysLeuLysLeuAspProPheLeuThrProTyThrLysIleAs	1144
QY	6765	CTCAAGATGAATTAAGACTTAAATATAGACATAAACCAATAAAACCCA-GAAGAAA	6823
Db	1144	nSerArgTrpLileLysAspLeuAsnValArgProLysThrLileLysThrLeuGluLys	1164
QY	6824	CCTAGGCAATACCATTCAGGATATGGACATGGCAAGACTTCATGACTAAACACCAAA	6883
Db	1164	nLeuGlyleThrileGlnAspLileGlyValGlyLysAspPheMetSerLysThrProLy	1184
QY	6884	AGCAATGCCACAAAGCCAAATAGACAACTGGATCTGATTAACATATAGACTCTGTG	6943
Db	1184	sAlaMetAlaThrLysAlaLysIleGlyLysTrpTyTrpLeuileLysLeuLysSerPheCy	1204
QY	6944	CACAGCAAAAACCTGTCATCAGAGTGAACAGCAACCTACAGAAATGGGAGAAATTT	7003
Db	1204	sThrAla-LysGluThrThrileArgValAsnArgGlnProThrThrTrpLileLysIleP	1224
QY	7004	TTGCAATCTATCGATCTGACAAAGCTTAATATCCAGAGATCTACGAAAGACTTAACAAA	7063
Db	1224	heAlaThrTyTrpSerAspLysGlyLeuileSerArgIleTyArgAsnGluLeuLysGlnI	1244
QY	7064	TTTACAGAAAAA-----AACAAACCCGTCAAATATATGGGCAAGATATAGCAGACACT	7119
Db	1244	leTyLysLysLysThrAsnAsnProileLysLysTrpAlaLysAspMetAsnArgHisP	1264
QY	7120	TCTCAAAAGAGACATTATGACCCCAACAAACATATGAAACAACTCATCATCTGG	7179
Db	1264	heSerLysGluAspLileTyArgAlaLysLysTyValLysLysCysSerSerLeuA	1284
QY	7180	TCGTTAGAGAAATCAAAACCAACCAAGTCACATCCATCTCTGCTAGTTAGAATGG	7239
Db	1284	laileArgGluMetGlnileLysThrMetArgTyThrHisLeuThrProValArgMeta	1304
QY	7240	TGATCACTAAAAGTCAGGAACCAACAAATCCTGGAGGATGTGGAGAAATA-----	7292
Db	1304	laileileLysLysSerGlyAsnAsnSerLysLysAsnLysAlaGlyIleMetProA	1324
QY	7293	-----GGAACA-----	7298
Db	1324	laaspPheLysLeuTyTyLysAlaMetGlyThrLysThrAlaTrpAspMetAspGluT	1344
QY	7299	-----C-----	7299
Db	1344	hrGluAsnHisAlaGlnGlnThrAsnThrArgThrGluLysGlnThrProHisValL	1364
QY	7300	TTTTCCACTGTGG-----TGG-----GAATGTAAT	7326
Db	1364	euThrHisLysTrpGluLeuAsnAsnGluAsnThrTrpThrGlnGlyGlyGluHisHis	1384
QY	7327	TAGTTCAACCAATTCGGAAGACACTGTGGAGATTCCTTAAGGATCTAGAACCAAGAAATAT	7386
Db	1384	hrProGlyProValArgGlyThrLeuGlyGlnLeu-----	1395
QY	7387	CATTGACCCAGCAATCCCATTAAGTATATACCCAAAGGAATATAAATCA-----	7439
Db	1396	-----ThrPheAlaArgLeuAsnLeuSerProLleSerTyTrpLysSerLeuGluA	1412
QY	7440	-----	7443
Db	1412	laSerAlaValThrSerValSerMetGluThrGlnIleArgLeuProLysMetArgPheT	1432
QY	7444	ATTATAAAGAC-----ACATGCACACAT-----ATGTTTATTGTCAG	7479
Db	1432	rpTyTrpGlnAspAspAlaGlyLeuileLysCysValGlyGluLeuLeuPhePheTyTC	1452
QY	7480	CACGTATCAACAATACGAAGACTTGGAAACCA-----CCC-----A	7515
Db	1452	ysLeuGluProPheGlnLysGluTrpTyTrpGlnLeuLeuSerValProLeuValGluPheG	1472
QY	7516	AATGTCCATCAGTATAGACTGGATAAAGAAACATGCACATATACACCATGAATACT	7575
Db	1472	lyCysGluSerValTrpSerTrpAlaPhePheGlyTrp-Asp-----	1485
QY	7576	ATGCAGCCATAAAAGGATGAGTTTCATGCTCTTTGCAGAGATATGGATGAAGCTGGAAC	7635
Db	1486	-----MetAspGluAlaGlyAsn	1491
QY	7636	CATCATTTCTCAGAACTAACACAGAACAGAAACCAACCAACCATGTTCTCCTACTGT	7695
Db	1492	HisHisSerGlnGlnThrAsnThrProThrGluAsnArgThrAlaHisLeuLeuThrHis	1511
QY	7696	AAGTGGAGTTGACAAATGACAGACATGGACAGGAGGAGGGGACACATCACACACCGT	7755
Db	1512	LysTrpGluLeuAsnAsnGluAsnThrTrpThrHisGlyLysHisHisHisThrProGly	1531
QY	7756	CCTGTTTCTGGTGC 7770	
Db	1532	ProValGlyGlyCys 1536	
RESULT 10			
ABG09152			
ID	ABG09152	standard; Protein; 1824 AA.	
XX			
AC	ABG09152;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	Novel human diagnostic protein #9143.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
FA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
DR	N-PSDB; AAS73339.		



xx New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

xx Claim 20; SEQ ID No 39511; 103pp; English.

xx The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

xx SQ Sequence 1824 AA;

#### Alignment Scores:

Pred. No.:	0	Length:	1824
Score:	4989.50	Matches:	1101
Percent Similarity:	67.01%	Conservative:	134
Best Local Similarity:	59.74%	Mismatches:	260
Query Match:	9.54%	Indels:	355
DB:	22	Gaps:	28

US-10-083-853B-2 (1-29921) x ABG09152 (1-1824)

QY	2490	GTAAATACAACTCTCCGAGCTAAAGGAGCATGTTCTTAACCCAA-----	2534
DB	131	IIeLeuAsnGlyGlnLysLeuGluAlaPheThrLeuLysThrGlyThrArgGlnGly	150
QY	2535	TCCAAGGAGCTAAAGACCTCTCAAAAAGTTAGATGATTCCTACTAGATATACGT	2594
DB	151	CysProLeuSerProLeuLeuPheAsnIleValLeuGluValAlaIleArg	170
QY	2595	GTAGAGAAACATAAATGACCTGATGGAGCTCAAAAACGCAAGACAACTTCATGAA	2654
DB	171	GlnGluLysGluIleLysGlyLe-----	178
QY	2655	GCATACACAGCTTCATAGCCAAATCGATCAAGCAGAAAGGATATCATGATTGAA	2714
DB	179	-----GlnLeuGlyLysGluGluValLysLeuSerLeuPheAla	191
QY	2715	GATCAAAAT-----	2723
DB	192	AspAspMetIleValTyrLeuGluAsnProIleIleSerAlaGlnAsnLeuLysLeu	211
QY	2723	-----	2723
DB	212	IleGlyAsnPheSerLysValSerGlyTyrThrIleAsnValGlnLysSerGlnAlaPhe	231
QY	2724	-----AATAAAGAAAGTCGAGAACAG-----	2750
DB	232	LeuTyrThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLeuProPheThrIle	251
QY	2751	ACAGAAAAGACTGAA-----AGAAACAACAAAGCCCAAGATATTGGACATATGTG	2807
DB	252	AlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspValLysAspPhe	271

QY	2808	AAA-----AGACCAAAATCTACATTTGTTGGTGTCCCCCAAGATGATGGGAGAAAT	2858
DB	272	LysGluAsnTyrLysProLeuLeuAsnGluIle-----LysGluAsp	285
QY	2859	GGAATCAAGTTGGAAACACTC-----	2882
DB	286	-LysAsnLysTyrLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMe	305
QY	2883	CAGGGTATTCACGAGGAATTTCCCATCTATCAGCGGAGGAGCAACATTCAAATTCAGG	2942
DB	305	talailLeuProLysArgThr-ProLeuLeuSerLeuProSerThrLeu---TyrGluG	324
QY	2943	AAATATGGAGA-----ACACCAATAAGATACTCTCTCGAGAGAA	2981
DB	324	luLeuTrpArgLeuArgSerTyrPheProLeuCysProGlnLeuTyrAsnMetGluGlnA	344
QY	2982	CAATCCCAAGACACATAATCTTCAGA-----TTCACCAAGTTGAAATGA	3026
DB	344	snThrThrAspTrpValIleTyrLysGluGlnCysIleTrpLeuThrAlaLeuGluThrG	364
QY	3027	AGGAAAAATGTTAAGGCGCAGCAGAGAGAAAGTTGGCTTACCCACAAGGAGCAACAA	3086
DB	364	lyGluSerLysSerThrAlaPro-----AlaPheGlyLysGlyTyrProM	379
QY	3087	TC-----AGACTAACA-----GCGGATCTCCCGGCGAGAA	3116
DB	379	etalaglutyrileIleIleTyrGlyArgLeuThrProHisMetAlaGlyTyrSerSerGlu	399
QY	3117	CCCTACACCCAGAGAGAGAGTGGGCCAATATTCACATCTTTAAAGAAAATAAATTTTC	3176
DB	399	hr-LysLeuProGluGluArgSerGly-----SerAsnIleCysCys	412
QY	3177	RAACCCAGATTTTCATATCCAGCCACACCAAGCTTCTTAAGTAGAGGAGAAATAAATCT	3236
DB	413	SerPro--IleSerAlaValLeuGlnProLeuLeuLeuLeuLeuProArgGlnThrGlySerG	432
QY	3237	CTACAGAG---AAGCAAAATGCTGACAGATTTTGTCCACCAGGCGCTCCCTTACAGAG	3293
DB	432	lyValAspLeuGlnGlnThrProThrAspLeuGlnLeuArgValLeuThrValArgArg-	451
QY	3294	CTCCTGAGGAGAGACCAACATCGGAAGAACAACTGGTACCAG-----CACTGCAAAAC	3350
DB	452	-----LysThrAsnLysGlnLysGlyHisProHisGlnAsnProIleCysThrS	468
QY	3351	ATCCCAAAATGTAAAGACCATTCATGCTATGAAGAAAGTGCATCACTAACGCGCAAAAT	3410
DB	468	erPro-----	469
QY	3411	AACCGAGTAGTGTCTAATATGCGAGGATCAATTCACACATAATAATATTAACTTAATG	3470
DB	470	-----SerSerLysThrLysGlySerAsnSerHisIleThrIleLeuThrLeuAsnV	487
QY	3471	TAAATGGGTAAATTCCTCCCAATTAAGACACAGACAGTGGCAATTCGATTAAGAGTCAAG	3530
DB	487	alAsnGlyLeuAsnAlaProIleLysArgHisArgLeuAlaAsnTrpIleLysSerGlnA	507
QY	3531	ACCCATCAGTGTGTATTTCAGGAGGCCCTCTCATGATGAAGAAAGACACATAGGCTCA	3590
DB	507	SerProSerValSerCysIleGlnGlnThrHisLeuThrCysArgAspThrHisArgIleL	527
QY	3591	AAATAAAGGATGGAGGAAGATTTACCAAGTAATGGAAACAAAAAAGGAGCGGG	3650
DB	527	ysileLysGlyTrpArgGluIleTyrGlnAlaAsnGlyLysGln---LysLysAlaGlyV	546
QY	3651	TTGCAATCTAGTCTCTGATAAACAGACTTTAAACCACAAAGATCAAAAGACAGAAAG	3710
DB	546	alalailLeuValSerAspLysThrAspPheLysProThrLysIleLysArgAspLysG	566
QY	3711	AGGGCCATACATAATGGTAAGGCATCAATGGAACAAGAGCTAACTATCTCTTAATA	3770
DB	566	luGlyHisTyrMetMetValLysGlySerIleGlnGlnGluLeuThrThrLeuAsnI	586



Qy	3771	TACATGCCCAATATACAGGAGCACCAGATTCTATAAGACGAAGTCTCTTAGAGACCTACAAA	3830	Db	906	----	906	----	906
Db	586	letyAlaProAsnThrGlyAlaProArgPheIleLysGlnValLeuArgAspLeuGlnA	606	Qy	4907	TAGGCCAAAGAGATACCGAGAGGGTGAGTTGGGAAGAGAGGTACAGCTGTTAGTCCAGAG	4966	Qy	4907
Qy	3831	GAGACTTTGACTCCACACAAATAATAGTGGAGTCTAAATAATAATAGACACTTTAAACA	3890	Db	906	----	906	----	906
Db	606	rgAspLeuAspSerHisThrLeuIleMetGly-----AspPheAsnT	620	Qy	4967	TAGGCCCTGCAAGACTGCATTCTCAAAACAATAGGCTCTAGATGTCCAGTAGATAACCTC	5026	Qy	4967
Qy	3891	CCCACGTGCCAATATTAGGCAGATCAATGACAGACAGAAAAATTAAACAAGGATATCCAGGAGT	3950	Db	906	----	906	----	906
Db	620	hrProLeuSerThrLeuAspArgSerThrArgGlnLysValAsnLysAspIleGlnAspL	640	Qy	5027	AAGGAGCCAGTCCAGGGAGTGATGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAAAGT	5086	Qy	5027
Qy	3951	TGAAGTGAAGTCTGGACCAACGAGCCTAATAGATATCTACAGAACTCCCCACCCCAAAAT	4010	Db	906	----	906	----	906
Db	640	euAsnSerAlaLeuHisGlnValAspLeuIleAspIleTyArgThrLeuHisProLys	660	Qy	5087	AACTTTGCCACATTTCTGTATTACAGATAAAACAGTTTGCTGTTTGATCAAGTAGCCTCCA	5146	Qy	5087
Qy	4011	CACAGAAATATACACTCTCTCAGCATCATCATACACCTATTATTAATTTGACCATGTAA	4070	Db	906	----	906	----	906
Db	660	erThrGluTyThrPhePheSerAlaLeuHisIleTySerLysIleAspHisIleTy	680	Qy	5147	GTGGAATGCTGAGTTGGTTCATGATCCCTTTGGCTTTTGGCTCCCAAAACACATACACC	5206	Qy	5147
Qy	4071	TTTTAAGTAAACACTCTCTCAGCAATGCAAAAAGAACAGAAATCCTAAACAAACAGTCTCT	4130	Db	907	-----	907	-----	907
Db	680	aiGlySerLysAlaLeuLeuSerLysTyThrGluIleIleThrAsnLysLeuS	700	Qy	5207	CTCTCAAGACTAAACCAGGAAGTCAAAATCCCTGAATATATACAGATACAAAGTTCTTAA	5266	Qy	5207
Qy	4131	CAGACTACAGTCAACTCTATTAGAACTCAGAATTAAAGAACTCACTCAAAATCACACAA	4190	Db	910	LeuProArgLeuAsnGlnGluValGluSerLeuAsnArgProIleThrGlySerGlu	929	Db	910
Db	700	erAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuThrGlnAsnArgSerT	720	Qy	5267	ATTGAAGCAGTAAATTGATAGCCTACCAACCAAAAAAGTCCAGGACCAGACCGATTCCACA	5326	Qy	5267
Qy	4191	CTACATGGAACGTGACAACTGCTCTGTAATGACTACTGGGTAAATAACAAATGAAGG	4250	Db	930	IleGluAlaIleIleAsnSerLeuProThrLysLysSerProGlyProAspGlyPheIle	949	Db	930
Db	720	hrThrTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyTrpValHisAsnLysMetLysA	740	Qy	5327	GCCAAATCTACAGAGGTACAAAGAGAGCTGGTACTATTCTCTGAAACTATTCCAA	5386	Qy	5327
Qy	4251	CAAAATAAAGATGTTCTTTGAACCAATCAGAACAAAGACACAATGTACCAGAACTCTCT	4310	Db	950	AlaGluPheTyGlnArgTyLysGluGluLeuValProPheLeuLeuLysLeuPheGln	969	Db	950
Db	740	laGluIleAsnThrLeuPheGluThrAsnGlnAsnLysAspThrThrTyThrGlnAsnLeu	760	Qy	5387	AAAAATA---GAAATATGGGAATCCTCCCTAACTCATTTTACGAGGCCAGCATCATCTGAT	5443	Qy	5387
Qy	4311	GGGGCATATTAAAGCAGTGTGAGAGGAAATTTATAGCACHAGATGCTCTACAAGAGAA	4370	Db	970	SerIleGlyLysGlu-GlyIleLeuProAsnSerPheTyGluAlaSerIleIleLeuIle	989	Db	970
Db	760	rpAspThrPheLysAlaValIcysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgL	780	Qy	5444	ACCAAAACCTAGCAGTGACACACACAAAAAGAGAAATTTTCAGGCCCATATCCCTGATGAA	5503	Qy	5444
Qy	4371	AGCAGGAAATATCTAAAATAGACACCTTAACATCACAAATTAAGAAGACTAGAGAAAG	4430	Db	989	eproLysProGlyArgAspThrThrLysLysAlaAsnPheArgProIleSerLeuMetAs	1009	Db	989
Db	780	ysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuGlyLysGlnG	800	Qy	5504	CATTGATGTGAAATCCTCAATAAATACTGGCAAAACCAATCCAGCAGCACATCAAAA	5563	Qy	5504
Qy	4431	AGCAACAAATTCAAAAGCTPAGCAGAGACAAACAAATACTAAGATCAGACGAGAAGTGA	4490	Db	1009	nIleAspAlaLysIleLeuAsnLysIleLeuAlaLysArgIleGlnGlnHisIleLysLys	1029	Db	1009
Db	800	luGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaLysLeuL	820	Qy	5564	GCTTATCTACCATGATCAAGTTGGCGTCACTCCCTGGGATGCAAGGCTGGTTCAAAATATG	5623	Qy	5564
Qy	4491	AGGAGATAGACACAAAAAGCCCTTCAATAAATAATCAATGAATCCAGGAGCTGGTTTTT	4550	Db	1029	sLeuIleHisAspGln-----	1035	Db	1029
Db	820	ysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpPhePhe	839	Qy	5624	CAAAATCAATTAATGTAGGCCATCACATAAACAGAACCAATGACAAAAACACATGATTAT	5683	Qy	5624
Qy	4551	GAAGAATCAGCAAAATAGACCA---CTAGACAGACTAATAAGAAAGAAAGAGAGAAG	4606	Db	1036	-----HisIleAsnLysAlaLysAspLysAsnHisMetIleIle	1048	Db	1036
Db	840	GluGlnIleAsnLysIleAspArgLeuLeuAlaArgLeuIleLysLysArgGluLys	859	Qy	5684	CTCAATAGATGCAGAAAAAGCCCTTTGCTCAAAAATTCACACGCCCTTCATCTCAAAAATTC	5743	Qy	5684
Qy	4607	AATCAAGAGATGCAATAAAAAATGATAAAGGGGATATCACACCCGATCCACAGAAATA	4666	Db	1048	eserIleAspAlaGluLysAlaPheAspLysIleGlnGlnProPheMetLeuLysThrLe	1068	Db	1048
Db	860	AsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIle	879	Qy	5744	CAGTAACTAGTATCGATGGAATGTATCTCAAAATAATAAGAGCTATTTTATAC-AAACC	5802	Qy	5744
Qy	4667	CAAACTATTATCAGAAATATTATAACACCTCTATGCAATAAATAGAAAAATCTAGAA	4726	Db	1068	uAsnLysLeuGlyIleAspGlyThrTyrrPheLysIleIleArgAlaIleTyrrAspLysPr	1088	Db	1068
Db	880	GlnThr							

```
Db 1128 gAlaIleArgGlnGluLysGluIleLysGlyIleGlnLeuLysGluLysGluValLysLe 1148
QY 5983 GTCTCTGTTGTCAGATGACATGTTGTATATTTAGAAAACCCCATCTCTCAGGCCAAAA 6042
Db 1148 uSerLeuPheAlaAspMetIleValTyLeuGluAsnProIleValSerAlaGlnAs 1168
QY 6043 ACTCCTTAGCTGATAGCAACTTCAGCAAAAGTCTCAGGACACAAATAATCAATGTGCAAAA 6102
Db 1168 nLeuLeuLysLeuIleSerAsnPheSerLysValSerGlyTyLeuLysIleAsnValGlnLy 1188
QY 6103 ATCACAAGCATCTTTATACGCCCAATAATAGACAAACAGAGACCAATCATGAGTGAAC 6162
Db 1188 sSerGlnAlaPheLeuTyThrAsnAsnArgGlnThrGluSerGlnIleMetSerGluLe 1208
QY 6163 CTCAATTCACAATTCCTACAAAGAGAATAAAATACCTAGGAATACAACTTACAAGGGACAC 6222
Db 1208 uProPheThrIleAlaSerLysArgIleLysTyLeuGlyIleGlnLeuThrArgSpVa 1228
QY 6223 GTAGGAACCTTCAAGGAGAACTTACAAACCACTGATCAAGGAATAAAGAGAGGACACAAA 6282
Db 1228 lLysAspLeuPheLysGluAsnTyLysProLeuLeuLysGluIleLysGluAspThrAs 1248
QY 6283 CAAATGGAAAAACATTCCTACGCTCACAGATAGTAAGAATCAT-----CAAAATGCCAT 6335
Db 1248 nLysTrpLysAsnIleProCysSerTrpValGlyArgIleAsnIleValLysMetAlaI 1268
QY 6336 ACTGCCCAAGTAATAATATAGATTTCAGTGTCTACCCCATCAAGCTACCATTTGACTTTCTT 6395
Db 1268 eLeuProLysValIleTyArgPheAsnAlaIleProIleLysLeuLeuIleThrPhePh 1288
QY 6396 CACAGAATTGGAAAAACAACCTTAAATTTTCATATGGAACCAACCAAAAAAGAGCCACAGAG 6455
Db 1288 eThrGluLeuGluTyThrThrLeuLysPheIleTrpAsnGln-LysArgAlaArgIleA 1308
QY 6456 CCAAGACAATCTTAAGCAAAAACAACAAAGCTGGAGGTATCATCTACCTGACTTAAAC 6515
Db 1308 lAlysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysL 1328
QY 6516 TATACATTAAGGCTACAGTAACCAAACTGCATGGTACTGGTACCAAAAACAGATATATAG 6575
Db 1328 eutyTyTyLysAlaThrValThrLysThrAlaTrpTyTrpTyGlnAsnArgAspIleA 1348
QY 6576 ACCAATGGACAGACAGACCTCAGAAATTAC-ACTGCAATCTACATCCATCCATGATCT 6634
Db 1348 spGlnTrpAspArgTrpGluProSerGluIleMetProHisIleTyAsnTyLeuIleP 1368
QY 6635 TTCACAAAACCTGACAAAACAACCAATGGAAGAAGGATCCCTTATTATAAATGTTGTT 6694
Db 1368 heAspLysProGluLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1388
QY 6695 GGAAGAACTGGCTAGCCATATGC---AGAAAGCTGAAACTGGATCCCTTCCCTTACACCTT 6751
Db 1388 rpGluAsnTrpLeuAlaIleCysArgArgLysLeuLysLeuAspProPheLeuThrPro 1408
QY 6752 ATACAAAAGTTAACTCAAGATGAATTAAGACTTAATAATAPACATAAAACCAATAAAAA 6811
Db 1408 yrThrLysIleAsnSerArgTrpIleLysAspLeuAsnValArgProLysThrIleLys 1428
QY 6812 CCAAGAGAAACCTTAGCAATACCATTCAGATATGGACATGGCGCAAGAGATTCATGA 6870
Db 1428 hrLeuGluGluAsnLeuGlyAsnThrIleGlnAspIleGlyMetGlyLysAspPheMets 1448
QY 6871 CTAAAAACCAAAAGCAATGGCAACAAAAGCCAAAATAGACAAGTGGGATCTGATTAAC 6930
Db 1448 erThrThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysL 1468
QY 6931 TATAGAGCTTCTGCACAGCAAAAAAAACTGTCTATCATCAGATGAACAGCAACCTACAGAA 6990
Db 1468 eulysSerPheCysThrAla-LysGluThrThrIleThrValAsnArgGlnProThrGlu 1487
QY 6991 TGGGAGAAATTTTCCATCTCATCTGATCTGACAAAGGCTAATATATCCAGAGATCTACGAA 7050
Db 1488 TrpGluLysIlePheThrIleTyProSerAspLysGlyLeuIleSerArgIleTyLys 1507
QY 7051 GAACCTTAACAAATTTTACAAAGAAAA-----AACAAACCCCTCAAAATATATGGCAAGGAT 7106
Db 1508 GluLeuLysGlnIleTyLysArgThrAsnAsnProIleAsnLysTrpValLysAsp 1527
QY 7107 ATGAGCAGACACTTCTCAAAAGAACACATTTATGACGCCCAACAAACATATGAAAAAAC 7166
Db 1528 MetAsnArgHisPheSerLysGluAspIleTyThrAlaLysArgHisMetLysLysCys 1547
QY 7167 TCATCATCAATGCTGTAGAGAAATGCAAAACAAACACACAGTCAGCATACCATCTCATG 7226
Db 1548 SerSerSerLeuAlaIleArgGluMetGlnIleLysThrThrLeuArgTyHisLeuIle 1567
QY 7227 CTAGTTAGAAATGTGTATCACTATAAAGTCAGGAAACAAACAACTCTCGAGAGGATGGGA 7286
Db 1568 ProValArgMetAlaIleLysLysSerGlyAsnAsnArgCysTrpArgGlyCysGly 1587
QY 7287 GAAATAGGAACACTTTTCCACTGTGTGGTGGGAATGTAATAGTTCAACCATTTGGGAAG 7346
Db 1588 GluIleGlyThrLeuLeuHisCysTrpTrpAsnCysLysLeuValGlnProLeuTrpLys 1607
QY 7347 ACAGTGTGGAGATTCCTTAAGGATAGAACCAAGAAATATCATTTGACCCAGCAATCCCA 7406
Db 1608 SerValTrpArgPheLeuArgAspLeuGluLeuGluIleProPheAspProAlaIlePro 1627
QY 7407 TTACTGAGTATATACCCAAAGGAATATAAATCATTTCTATTATAAGACACATGCACACAT 7466
Db 1628 LeuLeuGlyIleTyProLysAspTyLysSerCysCysTyLysAspThrCysThrArg 1647
QY 7467 ATGTTTATTCACACTGATCAATAGACAAAGACTTGGAACCAACCAATGTCATCA 7526
Db 1648 MetPheIleAlaAlaLeuPheThrIleAlaLysAlaTrpAsnGlnAlaLysCysProThr 1667
QY 7527 GTGATAGACTGGATAAGAAAAACATGCGACATATACACCATGAATACTATGACGCCATA 7586
Db 1668 MetIleAsnTrpIleLysLysMetTrpArgIleTyThrMetGluTyTyAlaAlaIle 1687
QY 7587 AAAAG-GATGAGTTCATGCTCTTTGTCAGAGATATGGATGAAGCTGGAAACCATCATTTCTC 7645
Db 1688 LysAsnAspGluPheMetSerPheValGlyThrTrpMetLysLeuGluIleIleLeu 1707
QY 7646 AGCAAACTACACAAAGACAGAAAAACCAACACACACATGTTCTCCTACTGTA 7696
Db 1708 SerLysLeuSerGlnGluGlnLysThrLysHisCysIleLeuSerLeuIle 1724
RESULT 11
ABG08447
ID -ABG08447 standard; Protein; 2921 AA.
AC ABG08447;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8438.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
```

DR WPI: 2001-639362/73.  
DR N-PSDB: AAS72634.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 38806; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2921 AA;

Alignment Scores:  
Pred. No.: 0 Length: 2921  
Score: 4902.00 Matches: 1065  
Percent Similarity: 63.19% Conservative: 111  
Best Local Similarity: 57.23% Mismatches: 223  
Query Match: 9.37% Indels: 469  
DB: 22 Gaps: 13

US-10-083-853B-2 (1-29921) x ABG08447 (1-2921)

QY 2343 GGGAAAAACAGTGCAGAACACTGAAATTCAAAACACAGAACTCTCTCTCAACCA 2402  
DB 2 GlyLysLysGlnSerArgLysThrGlyAsnSerLysLysLysGlnSerAlaSerProPro 21  
QY 2403 AAGATCACAACTCTCCAGCAGGAGGAAACAAACACAGATGGAGATGAGTTGAGGAA 2462  
DB 22 LysLysArgSerSerProAlaThrGluGlnSerThrThrGluAsnAspPheAspGlu 41  
QY 2463 TTGACAGAAAGTAGGCTTCAGAAAGTGGGTAAATACAAACTCTCCGAGCTAAAGAGCAT 2522  
DB 42 LeuArgGluGluGlyPheArg9-----SerAsnTyrSerGluLeuArgGluGlu 58  
QY 2523 GTTCTAACCCCAATGAGGAGAGCTTAAGAACCTTGAAAAAGGTTAGATGATTCCTAACT 2582  
DB 59 IleGlnThrLysGlyLysValGluAsnTyrGluLysSerLeuGluGluCysIleThr 78  
QY 2583 AGAATAATCAGTGTAGACAGAACATAATACCTGATGGAGCTGCAAAAACGCAAGACAA 2642  
DB 79 ArgIleThrAsnThrGluAsnCysLeuLysGluLeuMetGluProLysThrLysAlaArg 98  
QY 2643 GAACCTTCATGAGCATAACAGCTTCAATAGCCAAATCGATCAGCAGAAAGAGGATA 2702  
DB 99 GluLeuArgLysGluCysArgSerLeuArgSerArgCysTyrGlnLeuValGluArgVal 118  
QY 2703 TCAGTGATGAGATCAATTAATAAGAAAGTGTAGAGACAAAGATTACAGAAAAAGA 2762  
DB 119 SerAlaMetGluAspGluValAsn----- 126  
QY 2763 GTGAAGAACAACAACAGCTCCACAGAATTATGGGACTATGTGAAGAAGACCAATCTA 2822

DB 127 -----Glu-Met----- 128  
QY 2823 CATTTGATTGGTGTCCCCCAAGTATGGGGAGAAATGAATCAAGTTGAAAAACACTCTT 2882  
DB 128 ----- 128  
QY 2883 CAGGTATTATCCAGAGAAATTTCCCATCTATCAGGCGCAGGCCAACATTCAAATTCAGG 2942  
DB 128 ----- 128  
QY 2943 AAATATGAGAACACCACTAAAGATACCTCTCGAGAAACAATCCCAAGACACATAATCT 3002  
DB 128 ----- 128  
QY 3003 TCAGATTCCACCAAGTTGAAATGAAGAAATAATTTAAGGCGCAGCAGAGAAAGGTT 3062  
DB 128 ----- 128  
QY 3063 GGGTTACCCACAAAGGGAAGCAATCAGACTAACAGCGGATCTCCCGCAGAAACCCCTAC 3122  
DB 128 ----- 128  
QY 3123 AAGCCAGAAGAGAGTGAGGGCCAATATTCACATCTTTAAGAAAAATATTTTCAACCCA 3182  
DB 128 ----- 128  
QY 3183 GAATTTTCATATCCAGCCAAACCAAGCTTCCTAAGTGAAGAGAGAAATAAAATCCTCTACAG 3242  
DB 128 ----- 128  
QY 3243 AGAAGCAAAATGCTGACAGATTTTGTTCACCACCGGCTGCTTACAGAGCTCCTGAAG 3302  
DB 128 ----- 128  
QY 3303 GAAGCACCACAGATGGAAGGAAACAACTGGTACCACCACTGCATAAAACATCCCAATGT 3362  
DB 128 ----- 128  
QY 3363 AAAGACCATGATGCTGTGAAGAAAGTGCATCACTAACGGGCAAAATAAACAGCTAGTG 3422  
DB 128 ----- 128  
QY 3423 TCATAATGGCAGGATCAAAATTCACACATAATAATTAATTAACCTTAATTAATGGGCTAA 3482  
DB 128 ----- 128  
QY 3483 ATTCCTCCCAATTAAGACACAGACTGGCAAAATGGATAAAGAGTCAAGACCCATCAGTGT 3542  
DB 128 ----- 128  
QY 3543 GCTGTATTTCAGGAGGCCATCTCACATGAAAAGACACACATAGGCTCAAAATAAAGGAT 3602  
DB 129 -----LysAspThrHisArgLeuLysValLysGlyT 139  
QY 3603 GGAGAGAGATTACCAAGTAATGGAACAAAAAAGGAGGAGGCTGCAATCCCTAG 3662  
DB 139 rparGlyIleTyrGlnGlyAsnGlyLysGln---LysLysAlaGlyValAlaIleLeu 158  
QY 3663 TCTCTGATAAAACAGACTTTAAACCAACAAAGATCAAAAGAGACAAAGAGCCATTACA 3722  
DB 158 alSerAspLysThrAspPheLysProThrLysIleLysArgAspLysGluGlyHisTyrM 178  
QY 3723 TAATGGTAAAGCATCAATGGAACAAGAGCTAACTATCTTAATATATACATGACCCCA 3782  
DB 178 etMetValLysGlySerIleGlnGluGluLeuThrIleLeuAsnIleTyrAlaProA 198  
QY 3783 ATACAGAGCACCAGATTCATTAAGCAAGTCTTAGAGACCTACAAAGACACTTTGACT 3842  
DB 198 snThrGlyAlaProArgPheIleLysGlnValLeuSerAspLeuGlnArgAspLeuAsp 218  
QY 3843 CCCACACAATAATAGTGGGAGTCTAAATAATAAATAGACACTTTTAAACCCACCTGCCAA 3902  
DB 127 ----- 128

Db	218	erHisThrLeuIleMetGly-----AspPheAsnThrProLeuSerT	232
QY	3903	TATTAGCAGATCAATGACAGAGAAATTAACAAGGATATCCAGGAGTTGAAGTACGCTC	3962
Db	232	hrLeuAspArgSerThrArgGlnLysValAsnAsnThrGlnGluLeuAsnSerAlaL	252
QY	3963	TGACCAAGCGGACCTAATAGATATCTACAGAACTCCACCACCAATCAACAGAAATATA	4022
Db	252	euHisGlnAlaAspLeuIleAspIleThrArgThrLeuThrProLysSerThrGluThrT	272
QY	4023	CACCTCTCTCAGCATCACACTATTTAAATTTAAATTCACCATGTAATTTAAAGTAAAA	4082
Db	272	hrPhePheSerAlaProHisHisThrTySerLysIleAspHisIleLeuGlySerLysA	292
QY	4083	CACCTCTCAGCAATGCAAAAGAACAGAAATCTTAACAACACTCTCTCAGACTACAGTG	4142
Db	292	laLeuLeuSerLysCysLysArgThrGluIleIleThrAsnThrLeuSerAspHisSerA	312
QY	4143	CAATCTATTAGAACTCAGAATTAAGAACTCACTCAAAATCACAACTACATGGAAAC	4202
Db	312	laMetLysLeuGluLeuArgIleLysAsnLeuThrGlnAsnCysSerThrThrTrpLysL	332
QY	4203	TGAACAACCTGCTCCTGAATGACTACTGGTAAATAACAATAAGAGGCAAAATAAAGA	4262
Db	332	euAsnAsnLeuLeuAsnAspThrGlyValGlnAsnLysMetLysAlaGluIleLysM	352
QY	4263	TGTTCTTTGAAACCAATGAGAACACACACAAATGTACCAGATCTCTGGGCGATATTTA	4322
Db	352	etPhePheGluThrAsnGluAsnLysAspThrThrTyrgLnsnLeuThrPaspAlaPheL	372
QY	4323	AGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCCCTACAGAGAAAGCAGGAAATAT	4382
Db	372	ysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgS	392
QY	4383	CTAAATATACACCTTAACATCACATTAAGAACTAGAGAGAAAGAGCAACAAAT	4442
Db	392	erLysIleAspThrLeuThrSerGlnLeuLysGluLeuGluLysGlnGluInThrHisS	412
QY	4443	CAAAAGCTAGCAGAAGCAAGAAATAACTAAGATACAGACAGCAACTGAAGGAGATAGA	4502
Db	412	erLysAlaSerArgSerGlnGluIleThrLysIleArgAlaGluLeuLysGluIleGlnT	432
QY	4503	CACAAAAGCCCTTCAATTAATCAATGAATCCAGGAGCTGGTGTGTTTCAAAAGATCAGC	4562
Db	432	hrGlnLysThr-PheGlnLysIleAsnGluSerArgSerTrpPheGluArgIleAsn	451
QY	4563	AAAAT--AGACCACTAGACAGACTAATAAAGAAAGAGAGAGAGATCAAGAGAT	4618
Db	452	LysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAsp	471
QY	4619	GCAATAAAATGATAAGGGGATATCACACCCGATCCACAGAAATACAAACTATTATC	4678
Db	472	AlaIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrIle	491
QY	4679	AGAGAAATATTAACACCTCTATGCAATAAATACTAGAAATCTAGAGAAATGGATAAA	4738
Db	492	ArgGluThrTyThrLysHisLeuThrAlaAsnLysLeuGluAsnLeuGluMetAspThr	511
QY	4739	TTCTCTGCACACATATGACCTGTATGGACCTTGGGGGACAGACAAAGGGGTCATG	4798
Db	512	PheLeuAspThr-----	515
QY	4799	CAGAAATAAAGACAAAGACAAAGAGTATGTTGGAAGTAGGGGTGAGGGGCAACTTG	4858
Db	515	-----	515
QY	4859	CTCTTAATGGACAAGGCCCTGAGCTTTACACCACCTCTGTATTATTAGGCAAAAGAG	4918
Db	515	-----	515
QY	4919	ATAGCGAGAGGTGAGTTGGAAGAGAGGTACGCTGTAGTCCAGAGTAGGCCTGCAAG	4978
Db	515	-----	515
QY	4979	ACTGCATTCTCTCAACAATAGGCTCTAGATGTCCCAGTAGATAACCTCAAGGAGCCAGTG	5038
Db	515	-----	515
QY	5039	CCAGGAGTGATGGCCCTCAGCAACCTTCTTAGCGCAGCACAGAGTAAGTTTGGCCAC	5098
Db	515	-----	515
QY	5099	ATTCTGTATTACAGATAAACAGTTTGCTCTTTGATCAAGTAGCTCCAGTGAATCTGA	5158
Db	516	-----Cys---	516
QY	5159	GTTGTGTCATGATCCCTTTTGGCCCTTTTGGCTCCCAAAACACATACACCTCTCAAGACTA	5218
Db	517	-----ThrLeuProArgLeu	521
QY	5219	AACCAGGAAGAAGTCAAAATCCCTGATATATACAGTAACAAGTTCTAAAATTGAGACGTA	5278
Db	522	AsnGlnGluGluValGluSerProAsnArgProIleThrGlySerGluIleValAlaIle	541
QY	5279	ATTGATAGCTTACCAACCAAAAAAGTCCAGGACACAGCGGATTCACAGCCAAATTCAC	5338
Db	542	IleAsnSerLeuProThrLysLysSerProGlyProAsnGlyPheThrAlaGluPheThr	561
QY	5339	CAGAGTCAACAAGAGAGAGTGTACTTCTTCTGAAACTATTCCAAAAAATAGAA---	5395
Db	562	GlnArgTyLysGluGluLeuValProPheLeuLeuLysLeuPheGlnSerIleGluLys	581
QY	5396	ANTGGAAATCCCTCACTCAATTTTACAGGCGCAGCATCATCTCTGATACCAAAACCTAG	5455
Db	582	GluGlyIleLeuProAsnSerPheTyrgLluAlaSerIleThrLeuIleProLysAlaG1	601
QY	5456	CAGTGACACAAAGAGAGAAATTCAGGCCCATATCCCTGATCAACATTCATGCTGAA	5515
Db	601	yArgAspThrThrLysLysGluAsnPheArgProIleSerLeuMetCasnIleAspAla	621
QY	5516	AATCTCTCAATAAATACTGGCAACCAATCCAGCAGCATCATCAAAAAGCTTATCTACCA	5575
Db	621	sIleLeuAsnLysIleLeuAlaHisArgValGlnGlnHisIleLysLysLeuIleHisI	641
QY	5576	TGATCAAGTTGGGTCATCCCTGGGATCGAAGCTGGTTCACAAATATGCAAAATCAATAA	5635
Db	641	sAspGlnValGlyPheIleProGlyMetGlnCysTrpPheAsnIleCysLysSerIleAs	661
QY	5636	TGTAGCCCATCACATAAACAGAACCAATGACAAAAACCATGATTATCTCAATAGATGC	5695
Db	661	nValIleGlnHisIleAsnArgThrLysAspLysAsnHisMetIleIleSerIleAspAl	681
QY	5696	AGAAAAGGCTTTGTCTCAAAATTCACAGCCCTTCATGCTAAAATTTCTCAGTAAACTAGG	5755
Db	681	aGluLysAlaPheAspLysIleGlnGlnProPheIleLeuLysThrLeuAsnLysLeuG1	701
QY	5756	TATCGATGAATGTATCTCAAAATAATAAGAGCTATTATTAC-AAACCCACAGCCCAATAT	5814
Db	701	yIleHisGlyThrThrLeuLysIleIleArgAlaThrTyrsPlyProThrAspAsnI1	721
QY	5815	CATACGTAGTGGCAAAAACTGGAAGCTTCCTTTGAGAAGCTGGCAGACAGACAGATG	5874
Db	721	eLysLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrAspThrArgGlnGlyC	741
QY	5875	CCCTCTCTCACCACCTCTTATCAAGATATATTGGAAGTTCTGGCCAGGCGCAATCAGCA	5934
Db	741	sProLeuSerProLeuLeuPheAsnIleValLeuGluValLeuAlaArgAlaIleArgG1	761
QY	5935	ATAGAAGAAATAAAGGGTATTCAATAGAAAGAGAGAGATCATATTGTCTCTGTTTGC	5994
Db	761	nGluLysGluIleLysGlyIleGlnLysGluGluValLysLeuSerLeuPheAl	781
QY	5995	AGATGACATGTTGTATATTAGAAACCCCTCGTCTCAGGCCCAAAACTCCTTAAGCT	6054
Db	781	aAspMetIleValTyrrLeuGluAsnProIleValSerAlaGlnAsnLeuLeuLysLe	801

QY 6055 GATAGCAACATTCAGCAAGTCTCAGGACACAAAATCAATGTGCAAAAATCACAGCATT 6114  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerSerLeuAla 1180  
QY 801 uileSerAsnPheserLysValSerGlyTyrLysIleAsnValGlnLysSerGlnAlaPh 821  
QY 6115 CTTATACCCCAATATAGACAAAACAGAGAGCAAAATCATGAGTGAATCTCTCATCAAT 6174  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 821 eLeuTyrThrAsnLysArgGlnThrGlnSerGlnIleMetSerGluLeuProPheThrIl 841  
QY 6175 TGCTACAAAGAGATAAATACCTAGGAATACAACTTACAAGGACAGCTAGGAACTCTT 6234  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 841 eAlaSerLysArgGlyTyrLeuGlyIleGlnLeuThrArgaspValLysAspLeuPh 861  
QY 6235 CAAGGAGAACTACAAACCACTGATCAAGGAAATAAGAGAGACACAAACAAATGGAAAA 6294  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 861 eLysGluAsnTyrLysProLeuLeuLysGluIleLysGluAspThrAsnLysTyrLysAs 881  
QY 6295 CATTCATGCTCACAGATAGTAAGATCAT -----GAAATGCCATCTAGTCCCAAGT 6347  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 881 nileProCysSerValGlyArgIleSerPheValLysMetAlaIleLeuProLysVa 901  
QY 6348 AAATATAGATTGCTGCTACCCCACTCAAGTACCATTGCTTCTTTCACAGAAATGGA 6407  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 901 lileTyrArgPheAsnAlaIleProLysLeuProMetProPheProGluLeuGl 921  
QY 6408 AAAACAACTTTAAATTTATATATGGAACCAAAAAGAGCCACAGAGCCCAAGCAATCT 6467  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 921 uLysThrThrLeuLysPheIleTyrAsnGlnLysArgAlaArgIleAlaLysSerIleL 941  
QY 6468 TAACAAAAGAAACAACTGGAGGTATCATGCTACCTGCTTAACTTAACTATATTAAGG 6527  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 941 euSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLysLeuTyrTyrLysA 961  
QY 6528 CTACAGTAACCAAACTGCATGTGCTGCTACCAAAAACAGATATAGACCAATGGGAACA 6587  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 961 laThrValThrLysThrAlaTyrTyrTyrGlnAsnArgAspIleAspGlnTrpAsnA 981  
QY 6588 GAACAGAGACCTCAGAAATTAACCT -GCAATCTACATCCATCTGATCTTTGACAAACCTG 6646  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 981 rgThrGluProSerGluIleThrProHisIleTyrAsnTyrLeuIlePheAspLysProG 1001  
QY 6647 ACAAAAACAACAACTGGAAGAAGATCCCTATTATTAATGATGTTGGAAGAAGCTGC 6706  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1001 luLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCysTrpGluAsnTrpL 1021  
QY 6707 TAGCCATATGCAAGAAGCTGAACTGCTCCCTTACACCTTATACAAAGTTAACT 6766  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1021 euAlaIleCysArgLysLeuLysLeuAspProPheLeuThrProTyrThrLysIleAsn 1041  
QY 6767 CAAGATGAAATTAAGACTTAATATAGACATAAAACCATAAACCCA -GAACAAAACC 6825  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1041 erArgTrpIleLysAspLeuAsnValArgProLysThrIleLysThrLeuGluAsnL 1061  
QY 6826 TAGCAATACCATTCAGATATGACATGGCAAGACTTCTGACTTAAACACCAAAAG 6885  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1061 euGlyIleThrIleGlnAspIleGlyMetGlyLysAspPheMetSerLysThrGlnLysA 1081  
QY 6886 CAATGGCAACAAAACCCCAATAGCAAGTGGGATCTGATTAAACTATAGACTTCTGCA 6945  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1081 laMetAlaThrLysAlaThrIleAspLysTrpAspLeuIleLysLeuArgSerPheCyst 1101  
QY 6946 CAGCAAAAAAACTGTCATCAGATGAACAGCAACCTACAGAAATGGGAGAAATTTT 7005  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1101 hrAlaLysGluThrThrIleArgValAsnArgGlnProThrLysTrpGluLysIlePhe 1120  
QY 7006 GCAATCTCATGCTGCAAGAGGCTAATATCCAGAGATCTACGAAGAACTTAAACAAAT 7065  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1121 vallIeTyrSerSerAspLysGlyLeuIlePheArgIleTyrAsnGluLeuLysGlnIle 1140  
QY 7066 TACAAGAAAAA -----AACAACCCCGTCAAAATATGGCAAGGATATGACGACACTTC 7121  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 1141 TyrLysLysLysThrAsnAsnSerIleLysLysArgAlaLysAspMetAsnArgHisPhe 1160  
QY 7122 TCAAAAGAGACATTTATGACGCCCAACAACATATGAAAAAACCTCATCATCTGCTG 7181

Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1180  
QY 7182 GTTAGAGAAATGCAAAACCAACACAGTGCATACCATCTCTAGTGTAGTAGTG 7241  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1200  
QY 7242 ATCACTAAAAGTCAAGAAACCAAAATGCTGGAGAGATGTGGAGAAATAGGACACTT 7301  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1220  
QY 7302 TTCACATGTTGGTGGGAATGTAATAGTTCAACCATTTGGAAGACAGACAGTGTGAGATTC 7361  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1240  
QY 7362 CTTAAGGATCTAGAACCAAGAAATATCATTTGACCCAGCAATCCCATCTACTGATATATAC 7421  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1260  
QY 7422 CCAAGGAATATAAATCATTTCTATTATAAGACACATGCACACATATGTTTATTGACACA 7481  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1280  
QY 7482 CTGATCACATAGCAAAAGACTTGGACCAACCCCAAAATGTCATCATGATGATGATGATGATA 7541  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1300  
QY 7542 AAGAAAACATGGCCACATATACCATCAATGAATATCTATGAGCAATATAAAG -GATGAGTTC 7600  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1320  
QY 7601 ATGCTCTTTGAGAGATATGATGAAGTGGAAACCATCATTTCTCAGCAAACTAACACAA 7660  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1340  
QY 7661 GAACAAAACCAACCAACCATGTTCTCACTTGAAGTGGGAGTGTGACAAATGAGAGA 7720  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1354  
QY 7721 CATGGACACAGGAGAGGGG-----AACATCACACACACAGGCTCTGTTT 7762  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1374  
QY 7763 GTGGTGGGAGCTAGGAGAGGATAGCATTTAGGAGAAATACCTAACTAGTACGCGGT 7822  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1394  
QY 7823 TGATGGTGGCAGCAAGCCACCATGGCAGATGATACCTATGATGATGATGATGATGAT 7873  
Db |||||||SerLysGluAspIleTyrAlaAlaLysAsnHisMetLysLysCysSerSerLeuAla 1411  
RESULT 12  
ABG03983  
ID ABG03983 standard; Protein; 1691 AA.  
AC ABG03983;  
DT 13-FEB-2002 (first entry)  
DE Novel human diagnostic protein #3974.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS68170.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID No 34342; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1691 AA;

Alignment Scores:

Pred. No.: 0 Length: 1691  
Score: 4893.50 Matches: 1124  
Percent Similarity: 60.58% Conservative: 147  
Best Local Similarity: 53.57% Mismatches: 313  
Query Match: 9.35% Indels: 522  
DB: 22 Gaps: 32

US-10-083-853B-2 (1-29921) x ABG03983 (1-1691)

QY 1210 CATTCCCACTGAGTACCCAGTTCATCTC----- 1239  
DB 8 HisProProAlaGluThrProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27  
QY 1240 -----ACTGGGACTGGTTAGACATGGGTGCAGCCACCAGGAAGGTGAGCTGAA--- 1287  
DB 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47  
QY 1288 GCAGGTGGGTGCTCCCTCCAGCCGCGAGTGCACAGGGGTGGGGATCTCCTTCCCCCA 1347  
DB 48 GlnGlyThrAsnLysPro-----GlyIleGlnSerAspPro 59  
QY 1348 GCCAAGGGAAGCATGAGACGTGTACCAGAGGAAGTGTGTCACCTAGTCCAGATACTG 1407  
DB 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69  
QY 1408 CACTTTTCCATAGTCTTTGCAACTGGCAGACACGAGAGATTTCCCCAGTGCCTATGCCA 1467  
DB 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnHi 84  
QY 1468 C-----CAGGGCCCTGGGTTTCAACACAAACTGGGGGGCCATT----- 1507  
DB 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleGluPheIleTh 104

QY 1508 -----TGGACAGACACCGAGCTAGCCGACGACAGTTTATTTTTCATACCCAGTGG 1557  
DB 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124  
QY 1558 CGCCTGGAAATGCCAGCAGACAGAACCATTCACATCCAGGATCCAAAGTGTCTGGCTCAG 1617  
DB 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136  
QY 1618 T---GGGTCCACCCCATGGAGCCCACTA---GCTAAGATCCAC----- 1657  
DB 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysHi 156  
QY 1658 -----TGGCTTGAATCTCTCTCCGACGACGACGACGATC 1689  
DB 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176  
QY 1690 TGACATTGA-----CTGGG---ATGCTTGAGCTGG 1718  
DB 176 sGluValTyrSerMetArgThrMetGlyArgGlnProGlyThrAlaMetAspLeuAs 196  
QY 1719 TGAGGGGAGGGCGCTGCTGCCATTCCTGAGCTTGAGTAGCGGAGCGGTTTACCCCTCAA 1778  
DB 196 nAlaGlyGlySerIleCysAsnVal----- 204  
QY 1779 AGTGTAAACAAGACTACTGGGAAGTTTGAATGGGGCCGCCACCGCAGCTCAGCAAGCCG 1838  
DB 205 -----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr 222  
QY 1839 CTGTGGCAAACTGCTCTCTAGATTCTCTCTTTTGGGCGAGGTCTCTCTCAAGAAGAG 1898  
DB 222 gValGly-----AlatyrLeuAspIleArgG1 231  
QY 1899 CACAGCCCCAGTCAGGACTTATAGATAAAACCCCATCTCCCTG---GGACAGACA-- 1953  
DB 231 yTrpTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251  
QY 1954 -----ACCTGGGG-----AAGGGTGGCTGTGGGTGCAGCT 1985  
DB 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270  
QY 1986 TCTCAGACTTAAACATTCCTGCTGAGGCTCTGAAGAGAGACGCGGATCTCCAGCACCA 2045  
DB 271 -----As 271  
QY 2046 GCATTTCAGCTCTGATAAGGAGCAGCTGCTCCTCAAGTGGGTCCCTGACCCCATGTA 2105  
DB 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283  
QY 2106 TCCTGACTGGGAGACATCTCCCATTAGGGGCCAATAGACATTTTCATACAGGAGACAGGT 2165  
DB 284 -----I1 284  
QY 2166 CTGAGTGGACCTCCAGCAAACTCCAGACACCTGCAGCAGAGCGCGCTGACTGTAGAA 2225  
DB 284 eTrpValGlyGln----- 288  
QY 2226 GGAAGAGTAACAACAGAAAGGAATAGTATCAAC-----ATTAACAA 2267  
DB 289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305  
QY 2268 AAAGGACATCCACTCAGACACCCCATCTGAAGGTCAACACATCAAGAACCAAGGTAAA 2327  
DB 305 gLysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValas 325  
QY 2328 TAAACCAAAAGATGGGAAAAC-CAGTGCAGAAACACTGAAATTTCCAAAACCCAGAA 2386  
DB 325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345  
QY 2387 CTCTCTTCTCAACCAAGGATCACAACCTCTCCGACGAGGAGGACAAACACCATGGA 2446  
DB 345 rThrSerProThrLeuLysGluSerSerSerSerProAlaThrGluGlnSerTrpThrG1 365



QY	2447	GAATGAGTTTGAGGAATTGACAGAAAGTAGGGCTTCAGAGGTGGGTAATAACAAACCTCCTC	2506		Db	593	-----		596
Db	365	uAsnAspPheAspGluLeuArgGluGluGlyPheArg	382		QY	3586	GCTCAAAATAAAGGATGGAGGAAGATTACCAAGTAAATGGAAGAAAAAAGAAC	3645	-----AspThrHisar
QY	2507	CGAGCTAAGGAGCATGTTCTAACCCATGCAAGCAAGCTAAGAACCTTGAAAAAGGTT	2566		Db	596	gLeuLysIleLysGlyTrpArgLysIleTrpGlnAlaAsnGlyLysGln---	615	LysLysAl
Db	382	rgLLeuGlnGluGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe	402		QY	3646	AGGGTTGCAATCCCTAGTCTCTGATGATAAAGACAGCTTTAAACCAACAAGATCAAAAGAGA	3705	-----
QY	2567	AGATCAATTTGCTAACTAGATAATCAGTGTAGAGGAAGAACATAAATGACCTGATGGAGCT	2626		Db	615	aglyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs	635	-----
Db	402	uasplurprlieThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMeGluLe	422		QY	3706	CAAGAAGCCATTACATAATGCGTAAGGCATCAATGGACAAAGAGAGCTAACTATCCT	3765	-----
QY	2627	GAAGAACCAAGACAGAACTTCATGAAGCATACACAAAGCTTCATAGCCAAATCGATCA	2686		Db	635	physGluGlyHisTrpIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe	655	-----
Db	422	uLysThrLysAlaArgGluLeuCysAspGluArgThrSerLeuSerSerGlnCysAsnG	442		QY	3766	AAATATACATGACCCCAATACAGGACGCCAGATTTCATAAGCAAGTTCCTTAGACACCT	3825	-----
QY	2687	ACCAGAAAGAGATATCAGTCATTGACATCAAAATTAATAAAGAAAGTGAAGACAA	2746		Db	655	uasnlleTrpAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe	675	-----
Db	442	nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluGluLy	462		QY	3826	ACAAAGAGACTTTGACTCCCAACATAATAGTGGAGTCTAAATAAATAATAGACACTT	3885	-----
QY	2747	GATTACAGAAAAAGAGTCAAAAGAAACAAAGCCCTCCAAAGAAATTATGGGACTATGT	2806		Db	675	uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----	689	AspH
Db	462	sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyMe	482		QY	3886	TAACACCCACTGCCCAATATTAGGCAGATCAATGAGACAGAAATTAACAAGATATCCA	3945	-----
QY	2807	GAAGACCAATCTACATTTGATTGGTGTGCTCCCAAGAGTATGGGGAGAAATGGAATCA	2866		Db	689	easnthrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG	709	-----
Db	482	tlYsArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy	502		QY	3946	GGAGTTGAACCTGAGCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCCAACC	4005	-----
QY	2867	GTTCGAAACACTCTTCAGGGTATTATCCAGGAGAAATTTCCCAT-CTATCAGGCGAGC	2925		Db	709	nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTrpArgThrLeuHisPr	729	-----
Db	502	sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl	522		QY	4006	CAATCAACAGAAATATACACTCTTCAGCATCATCATACACTATTTTAAATATGCACCA	4065	-----
QY	2926	CAACATTCAAATTCAGGAAATATGGAGAACACCATTAAGATACTCTCCAGAGAACCAAT	2985		Db	729	oLysSerThrGluAsnThrPhePheSerAlaProHisIleThrTrpSerLysIleAspH	749	-----
Db	522	aAsnIleGlnIleGlnGluArgThrProGlnArgTySerLeuArgArgAlaTh	542		QY	4066	TCTAATTTTAACTAAACACTCTCTCAGCAAAATGCAAAAGAACAGAAATCTTAAACAACAG	4125	-----
QY	2986	CCCAAGACACATAATCTCAGATTCCACAGGTTGAAATGAAGGAAAAATTTAAGGC	3045		Db	749	sIleValGlySerLysAlaLeuLeuSerLysCysArgLysGluIleIleThrAsnCy	769	-----
Db	542	rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa	562		QY	4126	TCTCTCAGACTACAGTGAATCTATTAGAATCTAGAATTAAGAACTCCTCAAAATCA	4185	-----
QY	3046	AGCCAGAGAAAGGTGGTTTACCCACAAAGGGAAGCAATCAGACTAACAGCGATCT	3105		Db	769	sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy	789	-----
Db	562	lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe	582		QY	4186	CACACTACATGGAACATGAAACACTGCTCTCTGATGACTACTGGTAAATAACAANAAT	4245	-----
QY	3106	CCCGCAGAAACCTTACAGCCAGAGAGAGAGTGGGGCCCAATATTCCACATTCTTAAAGA	3165		Db	789	sSerThrIleTrpLysLeuAsnAsnLeuLeuAsnAspTyTrpValHisAsnGluMe	809	-----
Db	582	uLeuAlaGluThrLeuGlnAlaArgArgGln-----	592		QY	4246	GAAGCAAAATAAAGATGCTCTTTGAACCAATGAGAACAAAGACACAAATGTACCAGAA	4305	-----
QY	3166	AAATAATTTCAACCCAGAAATTTCAATATCCAGCCCAACCAAGCTTCTTAAGTGAAGAGA	3225		Db	809	tlYsAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrTyTrpGlnAs	829	-----
Db	592	-----	592		QY	4306	TCTCTGGGCGCATATTTAAAGCAGTGTGTAGAGGGAATTTATAGCACTAGATGCCCTACAA	4365	-----
QY	3226	AATAAAATCTCTACAGAGAGCAAAATGCTGACAGATTTTGTCCACCAGCGCTGCT	3285		Db	829	nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisG	849	-----
Db	592	-----	592		QY	4366	GAGAAGCAGGAAATATCTAAATAAGACACCTTACATCATCAATTTAAAGAACTAGAGAA	4425	-----
QY	3286	TACAAGAGCTCCTGAAGGAAGCACCACATGGAAGGAACAACTGGTACCAGCCACTGCA	3345		Db	849	yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLy	869	-----
Db	592	-----	592		QY	4426	GAAGAGCAAAACAAATTCAAAAGCTAGCAGAGAAAGAAATAACTTAACATCAGAGCAGA	4485	-----
QY	3346	AAAACATCCCAATTTGTAAGACCAATTGATGCTATGAAGAAGTGCATCACTAACGGGC	3405		Db	869	sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaG	889	-----
Db	592	-----	592		QY	4486	ACTGAAGGAGATAGACACAAAAAGCCCTTCAAATAAATCAATCAATCCAGGAGCTGT	4545	-----
QY	3406	AAAATAACCGCTAGTGTATTAATGGCAGGATCAAAATTCACACATAAATAATTAACCTT	3465		Db	889	uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrp	909	-----
Db	592	-----	592		QY	4546	TTTTTGAAGATCAGCAAAATAGACCACTAGACAGACTAATAAGAGAAAGAGAGAA	4605	-----
QY	3466	AAATGTAATGGCTAAATTTCCCAATTAAGACACAGACTGGCAAAATTTGGATAAGAG	3525		Db	909	Leu- - - - -Leu- - - - -ProLeuAlaArgLeuIleLysLysThrGluLy	922	-----
Db	592	-----	592		QY	4606	GAATCAAGAGATGCAATAAATAAAGGATATCAACCGGATATCACCAGCTCCACAGAAAT	4665	-----
QY	3526	TCAAGACCATCAGTGTCTGTATTTCAGAGGCCCATCTCACATGAAAAAGACACACATAG	3585						



Db 922 sAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAsnProThrAspIle 942  
QY 4666 ACAAACTATTATCAGAGAATATTATAAACAACCTCTATGCAATAATAAATACTAGAAAATCTAGA 4725  
Db 942 eGlnThrAlaIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuGI 962  
QY 4726 AGAAATGGGAAATTCCTGGACACATATGTAGCCTGTATGGACCTTGGGGGACAGAACAA 4785  
Db 962 uGluMetAspLysLeuLeu----- 968  
QY 4786 AAGGGGTGAATCAGAAATAAAGACAAAGACAAAGAGATGTTTGGNAGTAGGGTC 4845  
Db 968 ----- 968  
QY 4846 AGGGGGCAACTTGCCCTTAATGGACAAAGGGCCCTGAGCTTTACACCACCTCTGTATTTA 4905  
Db 968 ----- 968  
QY 4906 TTAGGCAAAGAGATAGCGAGAGGGTGAGTTGGAAGAAGAGTCAGCTGTAGGTCCAGA 4965  
Db 968 ----- 968  
QY 4966 GTAGGCCTCAAGACTGCATTCCTCAACAATAGGCTCTAGATGTCCAGTAGATAACCT 5025  
Db 968 ----- 968  
QY 5026 CAAGGAGCCAGTGCCAGGAGGTGATGGCCCTCAGCAAACTTCTAGGGCAGGCACAGAAG 5085  
Db 968 ----- 968  
QY 5086 TAAGTTGGCCCATCTGTATTTCACGATAAACAGTTGCTGTTGATCAAGTAGCCTCC 5145  
Db 968 ----- 968  
QY 5146 AGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCCTTTTGGCTCCCAAAACACATACAC 5205  
Db 969 -----AsnThrTyrThr 972  
QY 5206 CCTCTCAAGACTAAACACGAGAGAGTCAAAATCCCTGAATATACACAGTAACAGTTCTAA 5265  
Db 972 rIleProArgLeuAsnGlnGluValGluProArgAsnArgProIleThrGlySerGI 992  
QY 5266 AATTGAAGCAGTAATTGATAGCTACCAACCAAAAAAGTCCAGGACGAGCGGATTAC 5325  
Db 992 uIleGluAlaIleAsnArgLeuProThrLysLysCysProGlyProAspGlyPheTh 1012  
QY 5326 AGCCAAATCTACACAGGTACAAAGAGAAGCTGGTACTATTCTCTGAAACTATTCCA 5385  
Db 1012 rAlaGluPheTyrGlnArgTyrLysGln-----ValProPheLeuLysLeuPheGI 1030  
QY 5386 AAAAAATAGAA---AATGGGAATCCTCCCTAACTCAATTTTACGAGGCCAGCATCATCTGA 5442  
Db 1030 nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrGluAlaSerIleIleuI 1050  
QY 5443 TACCAAACTAGCAGTGACACACAAAGAGGAAATTCAGGCCCATATCCCTGATGA 5502  
Db 1050 leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta 1070  
QY 5503 ACATTGATGTGAAATCCCTCAATAAATACTGGCAAAACCAATCCAGCAGCACATCAAAA 5562  
Db 1070 snIleAspValLysIleLeuAsnLysIleLeuAlaLysGlnIleGlnHisIleLysL 1090  
QY 5563 AGCTTATCFACCATGATCAAGTTGGCGTCATCCCTGGGATGCAAGCTGGTTCAAAATAT 5622  
Db 1090 ysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA 1110  
QY 5623 GCAATCAATAATGTAGGCCATCACAATAACAGAACCAATGACAAAAACCATGATTA 5682  
Db 1110 rGlySerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnArgMetIleI 1130  
QY 5683 TCTCAATAGTCAGAAAAGGCTTTGTCAAATTCACAGCCCTTCATCGCTTAAAAATTC 5742  
Db 1130 leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL 1150

QY 5743 TCAGTAAACTAGGTATCGAATGTATCTCAAAATAATAAGAGCTATTATTATAC-AAAC 5801  
Db 1150 euAsnLysLeuGlyIleAspGlyThrTyrLeuLysThrIleArgAlaIleTyrAspLysP 1170  
QY 5802 CCACAGCCCAATATCATACTAGTGGGCAAAACAACTGGAAGCATTCCTTTGAGAACCTGGCA 5861  
Db 1170 roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyT 1190  
QY 5862 CAAGCAAGAGTCCCTCTCTCACCACTCTATTCAAGTACTATTGGAGTTCTGCGCA 5921  
Db 1190 hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuGluValLeuAla 1210  
QY 5922 GGCAATCAGGCAATAGAAAGTATTCAAAATAGAAAGAGAGAGAGAGTATCATAT 5981  
Db 1210 rgAlaIleSerGlnGluLysGluIleLysaspIleGlnLeuGlyLysGluGluValLysL 1230  
QY 5982 TGTCTCTGTTGGCAGATGACATGTTGTATATTTAGAAAACCCCTCTCTCAGGCCAAA 6041  
Db 1230 euSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnA 1250  
QY 6042 AACTCCTTAAGCTGATAAGCACTTCAGCAAAAGTCTCAGACACAAAAATCAATGTCAAA 6101  
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270  
QY 6102 AATCAACAAGCATTTTATACGCCAATAATAGACAAACAGAGAGCCCAAAATCATGAGTGAAC 6161  
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290  
QY 6162 TCTCATTCACAAATTCGTACAAAGAGATAAATACCTAGATATACAACTTACAGGAC 6221  
Db 1290 euProPheThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAsp 1310  
QY 6222 CGTAGAATCTTCAAGGAGAGACTCAAAACCACTGATCAAGGAAATAAGAGAGAGACACAA 6281  
Db 1310 alLysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThrA 1330  
QY 6282 ACAAAATGGAACAACTTCCATGCTCACAGATA-----GTAAGAATCATGAAAATGCC-A 6334  
Db 1330 snLysCysLysAsnValProCysSerTrpValGlyArgIleAsnIleMetLysMetAlaI 1350  
QY 6335 TACTGCCCAAGTAATATAGATTTCAGTGTCTACCCCTCAAGCTACCATTCACATTCTCT 6394  
Db 1350 leLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370  
QY 6395 TCACAAATTTGGAACAACTTTAAATTTTCATATGGAACCAACCAAGAGAGAGAGAGAG 6454  
Db 1370 heThrGluLeuGluLysThrThrLeuLysPheIleThrAsnGln-LysArgAlaCysIle 1389  
QY 6455 GCAAGACAATCTTAAGCAAAAGAACAAAGCTGGAGGTATCATGCTACTGACTTAA 6514  
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyGlyIleThrLeuProAspPheLys 1409  
QY 6515 CTATACTAAGGCTACAGTAACCAAACTGCATGGTGTACTGGTACCAAAACAGATATATA 6574  
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrThrTyrTrpTyrGlnHisArgAlaVal 1429  
QY 6575 GACCAATGCAACAGACAGAGACTCAGAAAATFAC-AC TGCAATCTACATCCATCTGATC 6633  
Db 1430 AspGlnTrpAsnArgThrGlyProSerGluIleMetProHisIleTyrAsnTyrLeuIle 1449  
QY 6634 TTTGACAAACCTGACAAAAACAGCAATGAAAAAGGATTCCTTATTAATAATGGTGT 6693  
Db 1450 PheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1469  
QY 6694 TGGAAAAACCTGGCTAGCCATATGCAGAAAGCTGAAAGCTCCCTTCCTTACACCTTAT 6753  
Db 1470 TrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnProPheLeuThrProTyr 1489  
QY 6754 ACAAAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATAAAACCAATAAAACC 6813  
Db 1490 ThrLysIleAsnSerArgTrpIleLysAspLeuHisValArgProLysThrIleLysThr 1509

```
QY 6814 CA-GAAGAAACCTTAGGCAATACCATTCAGGATATGGACATGGCGAAAGACATTCATGACT 6872
Db 1510 LeuGluGluAsnLeuGluYasnThrIleGlnValIleGlyMetGlyLysAspPheMetSer 1529
QY 6873 AAACACCAAAAGCAATGGCAACAAAGCCAAATAGACACAGTGGGATCTGATTAACATA 6932
Db 1530 LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeu 1549
QY 6933 TAGAGCTTCTGCACAGCAAAACAACTGTCATCAGAGTGACACAGCAACCTACAGAATG 6992
Db 1550 LysSerPheCysThrAla-LysGluSerThrIleArgValAsnArgGlnProThrGluTr 1569
QY 6993 GGAGAAATTTTGCATCTATCGATCTGCACAAAGGCTTAATATCCAGAGATCTACGAAGA 7052
Db 1569 pGluLysIlePheAlaThrTrpSerSerAspLysGlyGlnIleSerArgIleTyrAsnG1 1589
QY 7053 ACTTAACAATTTACAGAAAAA-----AACACCCCTCAAAATATATGGCAAGGATAT 7108
Db 1589 uLeuLysGlnIleTyrLysLysLysThrAsnAsnProIleLysLysTrpAlaLysAspAr 1609
QY 7109 GAGCAGACACTTCTCAAAAGAGACATTTATGCAGCCCAACAAACATATGAAAAAACCTC 7168
Db 1609 GAsnArgHisPheSerGluGluAspIleTyrAlaAlaLysLysHisMetLysLysCysSe 1629
QY 7169 ATCATCATTTGGTGTAGAGAAATGCAAAACAAACACAGTGCATACCATCTCATGCT 7228
Db 1629 rProSerLeuAlaIleArgGluValGlnIleLysThrThrMetArgTyrHisLeuThrPr 1649
QY 7229 AGTAGAATGTGATCACTAAAGTCAAGGAACACAA 7268
Db 1649 oValArgMetAlaIleIleLysLysSerGlyThrAsnArg 1662
```

## RESULT 13

ABG06058 .  
ID ABG06058 standard; Protein; 1691 AA.

```
XX AC ABG06058;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6049.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS70245.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 36417; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1691 AA;

Alignment Scores:  
Pred. No.: 0 Length: 1691  
Score: 4893.50 Matches: 1124  
Percent Similarity: 60.58% Conservative: 147  
Best Local Similarity: 53.57% Mismatches: 313  
Query Match: 9.35% Indels: 522  
DB: 22 Gaps: 32

US-10-083-853B-2 (1-29921) x ABG06058 (1-1691)

```
QY 1210 CATTCCCAACTGAGGTACCCAGTTCATCTC----- 1239
Db 8 HisProAlaGluTrpProArgHisValLeuLeuMetAlaAspSerLysAsnGlu 27
QY 1240 -----ACTGGGACTGGTTAGACATTTGGTGCAGCCCGACGGAAGTGAGCTGAA--- 1287
Db 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47
QY 1288 GCAGGTGGGGTGTCCCTCAGCCGCGAAGTGCAGGGGTGGGGGATCTCTTCCCCCA 1347
Db 48 GlnGlyTrpAsnLysPro-----GlyIleGlnSerAspPro 59
QY 1348 GCCAAGGAGACCATGAGACTGTACGAGGAGGAATGGTGCATCTAGTCCAGATACGT 1407
Db 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
QY 1408 CACTTTTCCCATAGTCTTTGCACTGGCAGCAGCAGGAGATTTCCCGCTGCTATGCCA 1467
Db 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnH1 84
QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAAACCTGGCGGCCATT----- 1507
Db 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleIleGluGluPheIleTh 104
QY 1508 -----TGGACAGACACCGAGCTAGCCGCGACAGCATTTATTTTTCATACCCAGTGG 1557
Db 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124
QY 1558 CGCCTGGAATGCCAGCAGACAGAACCATTCATCCAGGAGTCCCAAGTGGTCTGGCTCAG 1617
Db 124 rSer-----CysThrValLeuGlnValThrHisPheSe 136
QY 1618 T---GGGTCCACCCCATGGAGCCAGCTA---GCTAAGATCCAC----- 1657
Db 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysH1 156
QY 1658 -----TGCCTTGAATTTCTCTGCCAGCAGCAGTCC 1699
Db 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
QY 1690 TGAGATTGA-----CCTGGG---ATGCTTGAGCTGG 1718
Db 1718
```

```
Db 176 sGluValTyrSerMetArgThrMetGlyArgArgInProGlyThrAlaMetAspLeuAs 196
QY 1719 TGAGGGAGGGGGCTGCTGCATTTGCTGAGGCTTAGCGAGCGGGTTCCTCCCTCAA 1778
Db 196 nAlaGlyGlySerIleCysAsnVal-----204
QY 1779 AGTGTAAACAAGACTCTGGGAAGTTTGAATGGGGGCCGCCACCGCAGCTCAGCAAGCGCG 1838
Db 205 -----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr 222
QY 1839 CTGTGCGCAACTCCCTCTCTAGATTCTCTCTTTTGGCGAGGTCATCTCTCAAGAAGAGG 1898
Db 222 gValGly-----AlaTyrLeuAspIleArgG1 231
QY 1899 CAGCAGCCCGCAGTCAGGACTTATAGATAAAACCCCATCTCCCTG---GGACAGAGA-- 1953
Db 231 yTrpTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251
QY 1954 -----ACCTGGGG-----AAGGGTGTGCTGTGGTGCAGCT 1985
Db 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270
QY 1986 TCTCAGACTTAACATTCCTGCTGCGGCTCTGAAGAGCAGCGGATCTCCAGCACA 2045
Db 271 -----As 271
QY 2046 GCATTTGAGCTCTGATAAGGACGACGCTGCTCTCAAGTGGCTCCCTGACCCCATGTA 2105
Db 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly-----283
QY 2106 TCCTGACTGGAGACATCTCCCATATTAGGGGCCAATAGACATTTACAGGAGACAGGCT 2165
Db 284 -----11 284
QY 2166 CTGAGTGGACCTCCAGCAACTCCAGCAGACCTGCAGCAGAGCGGCTGACTGTAGAA 2225
Db 284 eTrpValGlyGln-----288
QY 2226 GGAAGAGTAAACAACGAAGAAAGTAATGATATCAAC-----ATTAACAA 2267
Db 289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305
QY 2268 AAAGGACATCCACTCAGACACCCCTCTGAAGTCAACAACTCAAGACCAAGGTA 2327
Db 305 gLysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValas 325
QY 2328 TAAACCAAAAAGATGGAAAAAC-CAGTCACAAACACTGAAATTCAAAACCCAGAA 2386
Db 325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345
QY 2387 CTCTCTCTCAACCAAGGATCACAACTCTCCGCCAGCAAGGAAACAAACCCAGATGGA 2446
Db 345 rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTrpThrG1 365
QY 2447 GAATGAGTTGAGAAATTGACAGAAGTAGGCTTCAGAAGTGGGTAAATACAAACTCCTC 2506
Db 365 uAsnAspPheAspGluLeuArgGluGluGlyPheArgArg-----SerAspTyrSe 382
QY 2507 CGAGCTAAAGGACGATCTTCAACCAATGTCAGGAAGCTAAGAACCTTCAAAAAGGTT 2566
Db 382 rGluLeuGlnGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe 402
QY 2567 AGATGAATTCCTAACTAGAATAATACGTGTAGAGAAGAACATAATATGACCTGTAGAGCT 2626
Db 402 uAspGluTrpIleThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMetGluLe 422
QY 2627 GAAACAGCAAGACAACTTCATGAAGCATACAAAGCTTCAATAGCCAAATCGATCA 2686
Db 422 uLysThrLysAlaArgGluLeuLysAspGluArgThrSerLeuSerSerGlnCysAsnG1 442
QY 2687 AGCAGAAGAAGATATCAGTATTGAAGTCAATTAATAAAGAAAGTGAAGACAA 2746
Db 442 nLeuGluGluArgValSerValMetGluAspGluMetAsnGlnIleLysGlnGluGly 462
```

```
2747 GATTACAGAAAAAGAGTGAAGAAACAAACAAAGCCTCCAAAGAATTATGGCACTATGT 2806
Db 462 sPheArgGluLysArgIleLysArgAsnGluInSerLeuGlnGlnIleTrpAspTyrMe 482
QY 2807 GAAAGACCAAAATCTCATTTGATTTGTTCTCCCAAGAGTATGGGAGAAATGAATCAA 2866
Db 482 tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy 502
QY 2867 GTTGGAAACACACTCTTCAGGTTATTCAGGAGAAATTTCCCAT-CTATCAGGGCAGGC 2925
Db 502 sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl 522
QY 2926 CAACATTCAAATTCAGAGAAATTCAGACACACCAATAAAGATATCTCTCCAGAGAGACAAT 2985
Db 522 aAsnIleGlnIleGlnGluIleGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh 542
QY 2986 CCAAGACACATATCTTCAGATTCCAAAGGTTGAAATGAAGAAATAATGTTAAGGGC 3045
Db 542 rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa 562
QY 3046 AGCCAGAGAGAAAGTTGGTTACCCACAAAGGGAAGCCCAATCAGACTAACACGGATCT 3105
Db 562 lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe 582
QY 3106 CCGCGCAGAAACCCCTACAAGCCAGAGAGAGTGGGGCCCAATATTCACATTTCTTAAGA 3165
Db 582 uLeuAlaGluThrLeuGlnAlaArgGln-----592
QY 3166 AAATAATTTTCAACCCAGAAATTCATATCCAGCCAAACCAAGCTTCTTAAGTGAAGAGA 3225
Db 592 -----592
QY 3226 AATAAAATCTCTACAGAGAAGCAATGCTGCACAGATTTTGTGCACCAGCAGCCTGCCT 3285
Db 592 -----592
QY 3286 TACAAGAGCTCTGAAAGAGCACCACATGGAAGGAACAACACTGGTACCAGCCACTGCA 3345
Db 592 -----592
QY 3346 AAAACATCCCAAAATTGTAAGACCATTGATGCTATGAGAAAGTGATCAACTAACGGGC 3405
Db 592 -----592
QY 3406 AAAATAACAGCTAGTGTCTAATATGCAGGATCAAAATTCACACATAATAATATAACCTT 3465
Db 592 -----592
QY 3466 AAATGTAAATGGCTAAATTCCTCAATTAAGAGACACAGACTGGCAAAATTTGGATAAGAG 3525
Db 592 -----592
QY 3526 TCAAGACCCTCAGTGTGCTGATTTCAGGAGGCCCTCTCATCATGAAAGACACACATAG 3585
Db 593 -----AspThrHisAr 596
QY 3586 GCTCAAAATAAAGGATGGAGGAAGATTACCAAGTAAATGGAAGAAACAAAAAAGC 3645
Db 596 gLeuLysIleLysGlyTyrArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAl 615
QY 3646 AGGGTGTGAATCTCTAGTCTCTGATAAACAGACTTTTAAACCAACAAAGATCAAAAGAGA 3705
Db 615 aGlyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs 635
QY 3706 CAAAGAGGCCATTACATATGTTAAGCATCAATGGAACAAGACAGCTAACTACTCT 3765
Db 635 pLysGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe 655
QY 3766 AAATATACATGACCCCAATACAGCAGCCACATTCATATAAAGCAAGTCTCTTAGACACCT 3825
Db 655 uAsnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe 675
```

QY	3826	ACAAAGAGACTTTGACTCCCAACACATAATACTAGGAGTCTAAATAATAATAGACACTT	3885	Db	968	-----	968
Db	675	uGlnArgAspLeuAspSerAsnThrIleIleThrGly	689	QY	4966	GTAGCCCTGCAAGACTGCATTCCTCAAACANTAGGCTCTAGATGTCCCGTAGATACCT	5025
QY	3886	TAACACCCACATGCCAATATAGGAGATCAATAGACAGAGAAATTAACAAGGATATCCA	3945	Db	968	-----	968
Db	689	easnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleG1	709	QY	5026	CAAGGAGCCAGTGGCCAGGAGTGATGGCCCTCAGCAAACCTTCTAGGCGAGCAGACAAG	5085
QY	3946	GGAGTTGAACCTGAGCTGGACCAAGCGGACCTAATAGATATCTACAGAACTCCGCCACC	4005	Db	968	-----	968
Db	709	ncluleuasnSerAlaLeuHisGlnAlaAspIleAspIleThrLeuHisPr	729	QY	5086	TAAGTTTGGCCACATTCGTATTACAGATAAACACTTTGCTGTTTGTATCAAGTACGCTCC	5145
QY	4006	CAATACACAGAAATATACACTCTCTCAGCATCATCATACACTATTTTAAATATGACCA	4065	Db	968	-----	968
Db	729	olysSerThrGluAsnThrPhePheSerAlaProHisHisThrTyrSerLysIleAspHi	749	QY	5146	AGTGAATGCTCAGTTGTCATGATCCCTTTTGGCCCTTTTGGCTCCCAAAACACATACAC	5205
QY	4066	TCTAATTTTAACTAAACACTCTCAGCAAAATGCAAAAGAACAGAAATCCTTAACAAACAG	4125	Db	969	-----	972
Db	749	sileValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy	769	QY	5206	CCTCTCAAGACTAAACACAGGAAGTCAAAATCCCTGAATATACAGTAACAAAGTTCTAA	5265
QY	4126	TCTCTCAGACTACAGTCAATCTATTAGAACTCAGAATTAAGAAACTCACTCAAAATCA	4185	Db	972	rileProArgLeuasnGlnGluValGluProArgAsnArgProIleThrGlySerG1	992
Db	769	sleuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy	789	QY	5266	AATTGAACGAGTAATTGATAGCTTACCACCAAAAAAGTCCAGGACGAGCGGATTCAC	5325
QY	4186	CACAACTACAGGAACCTGACAACTGCTCTCTGAATGACTACTGGTAAATAACAAAT	4245	Db	992	uileGluAlaIleIleAsnArgLeuProThrLysCysProGlyProAspGlyPheTh	1012
Db	789	sSerThrIleTrpLysLeuasnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMe	809	QY	5326	AGCCAAATCTACCAGAGGTACAAAGAGAGAGTGTACTATTCTCTCTGAACTATTCCA	5385
QY	4246	GAAGCAAAATAAAGATGTTCTTTGAACCAATGAGAACAAAGACACAATGTACCAGAA	4305	Db	1012	rAlaGluPheTyrGlnArgTyrLysGln-----ValProPheLeuLeuLysLeuPheG1	1030
Db	809	lLysAlaGluIleLysMeLeuPheGluThrAsnGluasnLysAspThrThrTyrGlnAs	829	QY	5386	AAAAATAGAA---AATGGAAATCCTCCCTCAATCTATTTTACGAGCCAGCATCTCTGA	5442
QY	4306	TCTCTGGGCGATATTAAAGCAGTGTGTAGAGGAAATTTATAGCACTAGATGCTCAAA	4365	Db	1030	nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrGluAlaSerIleIleLeuI	1050
Db	829	nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisG1	849	QY	5443	TACCAAACTAGCAGTGCACACACACAAAAAGAGGAAATTTTACGGCCCATATCCCTGATGA	5502
QY	4366	GAGAAAGCAGGAATATCTAAATAGACACCTTACATCACTCAATTAAGAACTAGAGAA	4425	Db	1050	leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta	1070
Db	849	yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLys	869	QY	5503	ACATTGATGTCAAAATCCTCAATAAAATCTGGCAAACTCCAGCAGCAGCATCAAAA	5562
QY	4426	GAAGAGCAAAATAATCAAAAGCTAGCAGAGCAAGCAAGAAATACTAAGATCAGAGCAGA	4485	Db	1070	snIleAspValLysIleLeuasnLysIleLeuAlaLysGlnIleGlnHisIleLysL	1090
Db	869	sglnGluGlnThrHisSerLysAlaSerArgGlnGluIleThrLysIleArgAlaG1	889	QY	5563	AGCTTATCTACCATGATCAAGTTGGCGTCTATCCCTGGATGCAAGCTGGTTTCAAAATAT	5622
QY	4486	ACTGAAGGAGATAGACACAAAAAGCCCTTCAATAAATCAATGAATCCAGGAGCTGGT	4545	Db	1090	ysLeuIleHisHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA	1110
Db	889	uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpp	909	QY	5623	GCAATCAATAATGTAGGCCCATCACATAAACAGACCAATGACAAAAACACATGATTA	5682
QY	4546	TTTTTTGAAAGATCAGCAAAATAGACCACTAGACAGACTAATAAGAGAAAGAGAGAA	4605	Db	1110	rgLysSerIleAsnValIleGlnHisIleAsnArgThrLysAspLysasnArgMetIleI	1130
Db	909	he-----Leu--ProLeuAlaArgLeuIleLysLysLysThrGluLys	922	QY	5683	TCTCAATAGATGCAAAAGGCCCTTTGTCAAAATTCACACGCCCTTCATGCTAAAAATTC	5742
QY	4606	GAATCAAAAGAGATGCAATTAATAAATGATAAAGGGGATATACACCGATCCACAGAAAT	4665	Db	1130	leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL	1150
Db	922	sasnGlnIleAspAlaIleLysasnAspLysGlyAspIleThrThrAsnProThrAspI1	942	QY	5743	TCAGTAACTAGTATCGATGGAATGATCTCAAAATAATAAGAGCTATTTATAC-AAAC	5801
QY	4666	ACAACTATTATCAGAGAAATTTATAACACCTCTATGCAATTAACCTAGAAAAATCTAGA	4725	Db	1150	euAsnLysLeuGlyIleAspGlyThrTyrLeuLysThrIleArgAlaIleTyrAspLysP	1170
Db	942	eGlnThrAlaIleArgGluTyrTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuG1	962	QY	5802	CCAGGCCAATATCATCTAGTGAATGGCAAAACTGGAGAGCATTCCTTTGAGAACTGGCA	5861
QY	4726	AGAAATGATAAATCTCTGGACACATATGTAGCCTGTATGACCTTTGGGGACAGACAA	4785	Db	1170	roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyT	1190
Db	962	uGluMetAspLysLeuLeu	968	QY	5862	CAAGACAAGGATGCTCTCTCAGCACTCTCTATTCAAGATACTATTGGAAGCTTCTGGCCA	5921
QY	4786	AAGGGGTGAATGCAGAAATAAAAGACAAAGACAAAGAGTATGTTTGGAAAGTAGGGTCT	4845	Db	1190	hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuGluValLeuAla	1210
Db	968	-----	968	QY	5922	GGGCAATCAGGCAATAGAAATAAAGGGTATTCAAAATAAGAGAGAGAGAGATCATAT	5981
QY	4846	AGGGGGCAACTTGCTCTAATGGACAAGGGCCCTGAGCTTTACACCACTCTCTGTATTTA	4905	Db	1210	rgAlaIleSerGlnGluLysGluIleLysAspIleGlnLeuGlyLysGluValLysL	1230
Db	968	-----	968	QY	5982	TGCTCTGTTGGCAGATGACATGTTTGTATATTTAGAAAAACCCCATCTCTCAGGCCAAA	6041
QY	4906	TTAGGCAAAAGAGATACGAGAGGGTGAGTTTGGAAAGAGAGGTCAGCTGTTAGTCCAGA	4965				

Db 1230 euSerLeuPhealaAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnA 1250  
QY 6042 AACTCCTTAAAGCTGATAGCAAACTTTCAGCAAAAGTCTCAGGACACAAAATCAATGTCACAA 6101  
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270  
QY 6102 AATCACAAGCATTCTATACGCCAATAATAGACAACAACAGAGAGCCAAATCATAGTAGAAC 6161  
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290  
QY 6162 TCTCATTCACAATTCTCACAAGAGCAATAAATACCTAGGATACAACTTACAAAGGACA 6221  
Db 1290 euProPheThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspV 1310  
QY 6222 CGTAGGAACCTTCAAGAGAGAACTACAAACCACTGATCAAGGAAATAGAGAGAGACAA 6281  
Db 1310 aLysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThrA 1330  
QY 6282 ACAATGGAAAACATTCCTCATCTCACAGATA-----GTAAGAATCATGAAAATGCC-A 6334  
Db 1330 snLysCysLysAsnValProCysSerTrpValClyArgIleAsnIleMetLysMetAlaI 1350  
QY 6335 TACTGCCAACTAAATATATACATTCAGTCTACCCCATCAAGCTACCATTCGACTTCT 6394  
Db 1350 leuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370  
QY 6395 TCACAGAAATGTGAAAAACAACATTTAAATTTTCATATGGAACCAAAAAAGAGCCACACAGA 6454  
Db 1370 heThrGluLeuGluLysThrThrLeuLysPheIleTrpAsnGln-LysArgAlaCysIle 1389  
QY 6455 GCCAAGACATCTTAAGCAAAAAGAACAAAGCTGGAGGTATCATGCTACCTGACTTAATAA 6514  
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLys 1409  
QY 6515 CTATACATAAGCTACAGTACCAACAACTGGTACTGGTACTGTACCAACACAGATATATA 6574  
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrThrTyrTyrTyrTyrGlnHisArgAlaVal 1429  
QY 6575 GACCAATGGAACAGACAGACCTCAGAAATTTAC-ACTGCAATCTACATCATCTGATC 6633  
Db 1430 AspGlnTrpAsnArgThrGlyProSerGluIleMetProHisIleTyrAsnTyrLeuIle 1449  
QY 6634 TTTGACAAACCTGCAAAAACAAAGCAATGGAAAAAGGATTCCTATTTAATAAATGGTGT 6693  
Db 1450 PheAspLysProAspLysAsnLysGlnTrpGlyLysAspSerLeuPheAsnLysTrpCys 1469  
QY 6694 TGGAAAACTGCTAGCCATATCCAGAAAGCTGAACCTGGATCCCTCTTACACCTTAT 6753  
Db 1470 TrpGluAsnTrpLeuAlaIleCysArgLysLeuLysLeuAsnProPheLeuThrProTyr 1489  
QY 6754 ACAAAGTTAACTCAAGATGAATTAAGACTTAAATATAAGACATATAAACCAACCAAAACC 6813  
Db 1490 ThrLysIleAsnSerArgTrpIleLysAspLeuHisValArgProLysThrIleLysThr 1509  
QY 6814 CA-GAAGAAAACCTAGGCAATACCATTCAGGATATGGACATGGGCAAGACTTCATGACT 6872  
Db 1510 LeuGluGluAsnLeuGlyAsnThrIleGlnValIleGlyMetGlyLysAspPheMetSer 1529  
QY 6873 AAACACAAAAGCAATGGCAACAAACCAAAATAGACAAGTGGATCTGATTAACATA 6932  
Db 1530 LysThrProLysAlaMetAlaThrLysAlaLysIleAspLysTrpAspLeuIleLysLeu 1549  
QY 6933 TAGAGCTTCTGCACAGCAAAAAAACTGTCAATCAGAGTGAACAGCAACCTACAGAAATG 6992  
Db 1550 LysSerPheCysThrAla-LysGluSerThrIleArgValAsnArgGlnProThrGluTyr 1569  
QY 6993 GGAGAAAATTTTGGCAATCTATCGATCTGCACAAAGGCTAATATCCAGAGATCTACGAAGA 7052  
Db 1569 pcuLysIlePheAlaThrTyrSerSerAspLysGlyGlnIleSerArgIleTyrAsnGln 1589  
QY 7053 ACTTAAACAATTTACAGAAAAA-----AACAAACCCCTCAAAATATGGCAAGCATAT 7108  
Db 1589 utLeuLysGlnIleTyrLysLysThrAsnAsnProIleLysLysTrpAlaLysAspAr 1609

QY 7109 GAGCAGACACTTCTCAAAAGAGAGACATTATGACCCAAACACATATGAAAAAACCTC 7168  
Db 1609 gAsnArgHisPheSerGluGluAspLeuTyrAlaAlaLysLysHisMetLysLysCysSe 1629  
QY 7169 ATCATCATTTGGTCGTAGAGAAATGAAACAAAACACAGATACCATCTCATGCT 7228  
Db 1629 rProSerLeuAlaIleArgGluValGlnIleLysThrMetArgTyrHisLeuThrPr 1649  
QY 7229 AGTTAGAATGGTGATCACTAAAAGCTCAGGAACACACAA 7268  
Db 1649 oValArgMetAlaIleLysLysSerGlyThrAsnArg 1662

RESULT 14  
ABG08641  
ID ABG08641 standard; Protein; 1691 AA.  
XX  
AC ABG08641;  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #8632.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS72828.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -  
XX  
PS Claim 20; SEQ ID No 39000; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

```
SQ Sequence 1691 AA;
Alignment Scores:
Pred. No.: 0 Length: 1691
Score: 4893.50 Matches: 1124
Percent Similarity: 60.58% Conservative: 147
Best Local Similarity: 53.57% Mismatches: 313
Query Match: 9.35% Indels: 522
DB: 22 Gaps: 32

US-10-083-853B-2 (1-29921) x ABG08641 (1-1691)
QY 1210 CATTCCCACTGAGGTACCCAGTTCATCTC----- 1239
D 8 HisProAlaGluTrpProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27
QY 1240 -----ACTGGGACTGGTTAGACATTGGGTGAGCCAGCCAGGAAGGTGAGCTCAA--- 1287
D 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47
QY 1288 GCAGGGTGGGTGTCCTCCCTCAGCCGCGAAGTGAAGGGGTGGGGGATCTCTTCCGCCCA 1347
D 48 GlnGlyTrpAsnLysPro-----GlyIleGlnSerAspPro 59
QY 1348 GCCAAGGAAGCCATGAGAGACTGTACAGAGGAAGTGGTGCACCTCTAGTCCAGATACGT 1407
D 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69
QY 1408 CACTTTTCCCATAGTCTTTGCACTGGCAGACAGGAGATTTCCCCAGTGCCTATGCCA 1467
D 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnH1 84
QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAACTGGCGCCCATTT----- 1507
D 84 svalmetglulysAlaValleuLysSerGlnAsnTrpAlaIleIleGluGluPheIleTh 104
QY 1508 -----TGGACAGACCCGAGCTAGCCGACAGCAGTATTTTTCATACCCAGTGG 1557
D 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124
QY 1558 CGCTTGAATCCAGCAAGACAGAACCATTCATCCAGGGATCCAAAGTGGTGGCTGCAG 1617
D 124 rSer-----CysThrValLeuGlnValThrHisPheSe 136
QY 1618 T---GGTCCCAACCCCATGAGCCAGCTA---GCTAAGATCCAC----- 1657
D 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysH1 156
QY 1658 -----TGGCTTCAATTCCTCTGCGCCAGCACAGCAGTC 1689
D 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176
QY 1690 TGAGATTGA-----CCTGGG---ATGCTTGGAGTGG 1718
D 176 sGluValTyrSerMetArgThrMetGlyArgGlnProGlyThrAlaMetAspLeuAs 196
QY 1719 TGAGGGAGGGCGTCTGCCATTCGTGAGGCTTGAGTAGCGGCGCGGTTTACCCCTCAA 1778
D 196 nAlaGlySerIleCysAsnVal----- 204
QY 1779 AGTGTAACAAGACTACTGGGAATTTGAATGGGGCGCCACCCAGCTCAGCAAGCGCG 1838
D 205 -----ProArgThrLeuLeuHisLeuThrGlyGluSerThrPheArgAspArgGlnAr 222
QY 1839 CTGTGGCAAACTGCTCTCTAGATTCCTCTCTTTTGGCAGGTCTATCTCGAAGAAAGG 1898
D 222 gValGly-----AlaTyrLeuAspIleArgG1 231
QY 1899 CAGCAGGCCAGTCAGGAGCTTATAGATAAAACCCCATCTCCCTG---GGACAGAGA--- 1953
D 231 yTrpTyrIleLeuValLeuValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251
QY 1954 -----ACCTGGGGG-----AAGGGTGGCTGTGGGTGCAGCT 1985
```

```
Db 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValLys-- 270
QY 1986 TCTCAGACTTAACATTCCTCGCTGGAGGCTCTCAAGAGACGACGGGATCTCCACGACA 2045
D 271 -----As 271
QY 2046 GCATTGAGCTCTGATAAGGACAGCGCTGCTCTCAAGTGGGTCCCTGACCCCATGTA 2105
D 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283
QY 2106 TCCTGACTGGGAGACATCTCCATTAGGGGCCAATAGACATTTTCATACAGGAGACAGGT 2165
D 284 ----- 284
QY 2166 CTGGAGTGGACCTCCACAAACTCCAGCAGACCTGCAGACAGCGGCTGACTGTAGAA 2225
D 284 eTrpValGlyGln----- 288
QY 2226 GGAAGAACTAACAAACAGAAAGGAATAGTATCAAC-----ATTAACAA 2267
D 289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305
QY 2268 AAGGACATCCACTCAGACAGCCCATCTGAAGGTCAACACATCAAGACCAAGGTAAA 2327
D 305 gLysAspIleHisIleLysAsnProSerValCysHisHisGlnArgProLysValAs 325
QY 2328 TAAACCAAAAGATGGAAAAAC-CAGTGCAGAAACATGAAATTTCCAAAACACGAA 2386
D 325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345
QY 2387 CTCCTCTCTCAACAAAGGATCAACTCCTCCCGCAGAGGGAACAAACACAGATGGA 2446
D 345 rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTrpThrG1 365
QY 2447 GAATGAGTTTCAGAAATGACAGAAAGTAGGCTTCAGAGGTGGGTAAATAACAACCTCCTC 2506
D 365 uAsnAspPheAspGluLeuArgGluGlyPheArgArg-----SerAspTyrSe 382
QY 2507 CGACTTAAAGGAGCATGTTCTTAACCCCAATGCAAGAAAGCTTAAGAACCTTGAANAAGGTT 2566
D 382 rGluLeuGlnGluGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe 402
QY 2567 AGATGAATTCGTACTAGATAATCATCTAGTAGAGAAACATAAATACCTGATGGAGCT 2626
D 402 uAspGluTrpIleThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMetGluLe 422
QY 2627 GAAAAACGACAGACAACTTTCATGAAGCATACACAAAGCTTCAATAGCCAAATCGATCA 2686
D 422 uLysThrLysAlaArgGluLeuCysAspGluArgThrSerLeuSerSerGlnCysAsnG1 442
QY 2687 AGCAGAAAGAAAGGATATCAGTGTGAAGATCAAAATTAATAAAGAAAGTGAGAAGACA 2746
D 442 nLeuGluGluArgValSerValMetGluAspGluMetAsnGluIleLysGlnGluLys 462
QY 2747 GATTACAGAAAAGAGCTGAAAACAAACAAAGCCTCCCAAAATATATGGGACTATGT 2806
D 462 sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTrpAspTyrMe 482
QY 2807 GAAAGACCAACTACATTTGATTGTTGTCCTCCCAAGTGCATGGGAGATGGAATCAAA 2866
D 482 tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy 502
QY 2867 GTTGGAAAAACCTCTTCAGGGTATTATCCAGGAGAAATTTCCCAT-CTATCAGGCGCAGGC 2925
D 502 sLeuGluAsnThrLeuGlnAspIleIleGlnGluAsnPheProAsnLeuAlaArgGlnAl 522
QY 2926 CAACATTTCAATTCAGGAAATATGGAGAACACACCAATAAAGATCTCTCGAGAGAACAAT 2985
D 522 aAsnIleGlnIleGlnGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh 542
QY 2986 CCCAAGACACATATCTTCAGATTCACCAAGGTTGAAATGAGGAAAAAATGTTAAGGC 3045
```



Db	542	rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgVa	562
QY	3046	AGCCAGAGAAAGGTTGGTTACCCACAAAGGGAAGCCAAATCAGACTAACAGCGGATCT	3105
Db	562	lalaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe	582
QY	3106	CCCGCAGAAACCCCTACAGCCAGAGAGAGTGAGGGCCAAATATTCACACATCTTTAAAGA	3165
Db	582	uLeuAlaGluThrLeuGlnAlaArgGln	592
QY	3166	AAATAATTTTCAACCCAGAAATTCATATCCAGCCAAACCAAGCTTCCTAAGTGAAGGAGA	3225
Db	592	-----	592
QY	3226	ATAATAATCCTCTACAGAGAAGCAAAATGCTGACAGATTTTGTGCACCAACGCGCTGCCT	3285
Db	592	-----	592
QY	3286	TACAAGAGCTCCTGAAGGAAGCACCACCAATGTAAGGAAGCAAACTGGTACCACCCACTGCA	3345
Db	592	-----	592
QY	3346	AAACATCCCAAAATGTAAGACCATTGATGCTATGAAGAAAGTGCCATCACTAACGGGC	3405
Db	592	-----	592
QY	3406	AAAAAACCAGCTAGTGTCTAATAATGCCAGGATCAAAATTCACACATAATAATTAACCTT	3465
Db	592	-----	592
QY	3466	AAATGTAATGGCTAAATTCCTCAATTAAGACACAGACTGGCAAAATGGATAAAGAG	3525
Db	592	-----	592
QY	3526	TCAAGACCCTCAGTGTGTGTATTTCAGAGGCCCATCTCACATGAAAAGACACACATAG	3585
Db	593	-----	596
QY	3586	GCTCAAAATAAAGGGATGGAGGAAGATTTACCAAGTAATGGAACAAAAAAGG	3645
Db	596	gLeuLysIleLysGlyTrpArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAl	615
QY	3646	AGGGTTGCAATCCTAGTCTCTGATAAAGACAGCTTTAAACCAACAAAGATCAAAAGAGA	3705
Db	615	agLyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs	635
QY	3706	CAAGAAGGCCATTACATAATGGTAAAGCATCAATGGACAGAGAGCTACTATCCT	3765
Db	635	physGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluLeuThrIleLe	655
QY	3766	AAATATACATGCCCAATACAGGACGCCAGATTCATAAAGCAAGTTCTTAGAGACCT	3825
Db	655	uasnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuArgAspLe	675
QY	3826	ACAAGAGACTTTGACTCCCAACACATAATAGTGGGAGTCTAAATAATAAATAGACACTT	3885
Db	675	uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspPh	689
QY	3886	TACACCCCTGCCCAATATTAGGCAGATCAATGAGACAGAAATAAAGAGATATCCA	3945
Db	689	easnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleGl	709
QY	3946	GGAGTTGAAGCTGAGCTCGACCAAGCGACCTAATAGATATCTACAGAACTCCCAACCC	4005
Db	709	nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisPr	729
QY	4006	CAAATCAACAGAAATATACACTCTCTCAGCATCATCACTTACCTATTTTAAATATGACCA	4065
Db	729	oLysSerThrGluAsnThrPhePheSerAlaProHisHisThrTyrSerLysIleAspHi	749
QY	4066	TGTAATTTTAAATAAACACTCTCTCAGCAAAATGCAAGAAACAGAAATCCTTACAAACAG	4125
Db	749	sileValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy	769
QY	4126	TCTCTCAGACTACAGTGCATCTATTATTAGAACTCAGAATTAAAGAACTCACTCAAAATCA	4185
Db	769	sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy	789
QY	4186	CACAACACTAGTGAAGAACTGAACACCTGCTCTCTGAATGACTACTGGTGAATAACAAAT	4245
Db	789	sSerThrIleTrpLysLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMe	809
QY	4246	GAAGCAAAATAAAGATCTTTTGAACCAATGAGAAACAAGACACAACTGTPACCAGAA	4305
Db	809	tlYsAlaGluIleLysMetLeuPheGluThrAsnGlnLysAspThrThrTyrGlnAs	829
QY	4306	TCTCTGGGCATATTATTAAGACGTGTAGAGGAAATTTATAGCACTAGATGCCTACAA	4365
Db	829	nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisGl	849
QY	4366	GAGAAAGCAGGAAATATCTAAATAGACACCTTAAACATCAATCAATTAAGAACTAGAGAA	4425
Db	849	yArgLysGlnGluArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLy	869
QY	4426	GAAGAGCAAAACAATTCAAAGCTAGCAGACAGACAAAGAAATAACTAAAGATCAGACAGA	4485
Db	869	sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaGl	889
QY	4486	ACTGAAGGAGATAGACACACAAAAGCCCTTCAAATAAATCAATGAATCCAGGAGCTGGT	4545
Db	889	uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleasnGluSerArgSerTrp	909
QY	4546	TTTTTGAAGATCAGCAAAATAGACCTAGACACTAGACACTAAATAAGAAAGAAAGAGAGA	4605
Db	909	he-----Leu--ProLeuAlaArgLeuIleLysLysLysThrGluLy	922
QY	4606	GAATCAAGAGATGCAATAAATAATGATAAAGGGATATCACACCGATCCCAACAGAAAT	4665
Db	922	sasnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrThrAsnProThrAspIl	942
QY	4666	ACAAACTATATCAGAGAAATATTATAACACCTCTATGCAAAATAAACTAGAAAATCTAGA	4725
Db	942	eGlnThrAlaIleArgGluTyrTyrLysHisLeuTyrAlaasnLysLeuGluAsnLeuGl	962
QY	4726	AGAAATGGATAAATTCCTCGACACATATGTAGCTCTATGGACCTTGGGGGACAGACAA	4785
Db	962	uGluMetAspLysLeuLeu-----	968
QY	4786	AAGGGGTCGAATGCAGAAATAAAGACAAAGACAAAGAGATGTGTTTGAAGTAGGGGTC	4845
Db	968	-----	968
QY	4846	AGGGGCAACTTGCCTCTAATGGACAAGGGCCCTGAGCTTTACACCCCTCTGTATTTA	4905
Db	968	-----	968
QY	4906	TTAGGCAAAAGAGATACCGAGAGGGTGAGTTGGAAGAAGAGGTCACCTGTTAGGTCAGA	4965
Db	968	-----	968
QY	4966	GTAGGCTGCAAGACTGCATTCTCTCAACAATAAGGCTCTAGATGTCCAGTAGATAACCT	5025
Db	968	-----	968
QY	5026	CAAGGACCAAGTCCAGGGAGTGATGGCCCTCAGCAAAACCTCTTAGGGCAGGCACAGAG	5085
Db	968	-----	968
QY	5086	TAAGTTTGGCCACATTTCTGTATTCAGATAAACAGTTTGTGTTTGTATCAAGTAGGCTCC	5145
Db	968	-----	968
QY	5146	AGTGAATGCTGAGTTGGTTCATCATCCCTTGGCCCTTTTGGCTCCCAACACACATACAC	5205
Db	969	-----AsnThrTyrTh	972





DT 13-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #10793.  
 XX Human; chromosome mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
 KW OS Homo sapiens.  
 XX W0200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-0508631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB; AAS74989.  
 XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity  
 XX Claim 20; SEQ ID No 41161; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1691 AA:  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 1691  
 Score: 4893.50 Matches: 1124  
 Percent Similarity: 60.58% Conservative: 147  
 Best Local Similarity: 53.57% Mismatches: 313  
 Query Match: 9.35% Indels: 522  
 DB: 22 Gaps: 32  
 US-10-083-853B-2 (1-29921) x ABG10802 (1-1691)  
 QY 1210 CATTCCCACTGAGGTACCCAGTTCATCTC----- 1239  
 DB 8 HisProAlaGluTrpProArgHisValLeuLeuLeuMetAlaAspSerLysAsnGlu 27  
 QY 1240 -----ACTGGGACTGTTAGACATTGGGTGCGCCACCGAAGGTGACGTGAA--- 1287  
 DB 28 LeuGlnAsnMetAsnThrGlyGlnThrCysGlyThrArgGlnThrLeuLysSerLysLeu 47

QY 1288 GCAGGTGGGTGTCCCTCAGCCGGAAGTGCAGGGGTGGGATCTCTCTCCCCCA 1347  
 DB 48 GlnGlyTrpAsnLysPro-----GlyLeuGlnSerAspPro 59  
 QY 1348 GCCAAGGGAAGCATGAGACAGTGTACAGGAGGAATGTGCTACTCTAGTCCAGACTG 1407  
 DB 60 GlySerGlnGlyHisProArg-CysLeuLys----- 69  
 QY 1408 CACTTTTCCATAGTCTTTTGCACACTGGCAGACCAGGAGATTTCCCCAGTGCCTATGCCA 1467  
 DB 70 -ThrPheProLeuValAlaGlnSerSer-----ThrGlnGluArgAsnH1 84  
 QY 1468 C-----CAGGGCCCTGGGTTTCAAGCACAAAACTGGGGCCCATTT----- 1507  
 DB 84 sValMetGluLysAlaValLeuLysSerGlnAsnTrpAlaIleGluGluPheIleTh 104  
 QY 1508 -----TGGACAGACACCGAGCTAGCCGAGCAGGTATTATTTTATACCCAGTGG 1557  
 DB 104 rLysIleGlyTrpPheProAspTyrSerGlyLysValLysTyrLeuSerSerProLysSe 124  
 QY 1558 CGCTCGAATGCCAGCAGACAGACCACTTCACTCCAGGATCCAGTGTCTGGCTCAG 1617  
 DB 124 rSer-----CysThrThrValLeuGlnValThrHisPheSe 136  
 QY 1618 T---GGGTCCACCCCATGGAGCCAGCTA---GCTAAGATCCAC----- 1657  
 DB 136 rLeuPhePheHisProAsnProAspGluLeuValSerLysIleLysValTrpSerLysH1 156  
 QY 1658 -----TGGCTTGAATTCCTCTCCGCCAGCAGCAGCAGTC 1689  
 DB 156 sArgLeuTyrGlnAsnAsnSerGluAlaPheLeuGluValGlnIleProGluProLysCy 176  
 QY 1690 TGAGATTGA-----CCTGGG---ATGCTCAGCTGG 1718  
 DB 176 sGluValTyrSerMetArgThrMetGlyArgArgGlnProGlyThrAlaMetAspLeuAs 196  
 QY 1719 TGAGGGAGGGGGCTGCTGCATTGCTCAGGCTTGTAGTGGCAGGCGGTTTACCTCAA 1778  
 DB 196 nAlaGlyGlySerIleCysAsnVal----- 204  
 QY 1779 AGTGTAAACAAGCTACTGGGAAGTTGAATGGGGGCCCCAGCAGCTCAGCAAGCCG 1838  
 DB 205 -----ProArgThrLeuLeuHisLeuThrGlyGlyLeuSerThrPheArgAspArgGlnAr 222  
 QY 1839 CTGTGCAAACTGCCCTCTCTAGATTCTCTCTTTTGGCAGGAGTCTCTGAAAGAAAGG 1898  
 DB 222 gValGly-----AlaTyrLeuAspIleArgG1 231  
 QY 1899 CAGCAGCCCGCTCAGGACTTATATATAAACCCCATCTCCCTG---GGACAGAGA-- 1953  
 DB 231 yTrpTyrIleLeuValLeuValValAspLysGluLysLeuGlyIleProGlyGlnLysG1 251  
 QY 1954 -----ACCTGGGG---AAGGGGTGGCTGTGGGTGCAGCT 1985  
 DB 251 nAsnSerSerLysIleLeuPheGlySerGluLysLeuLysGlnTrpValTyrValLys-- 270  
 QY 1986 TCTCAGACTTAAACATTTCTGCTGAGGAGTCTGAAGAGAGCAGCGGATCTCCAGCACA 2045  
 DB 271 -----AS 271  
 QY 2046 GCATTTGAGCTCTGATAAGGAGACAGCTGCCTCCTCAAGTGGGTCCCTGACCCCATGTA 2105  
 DB 271 nHisAlaLysGluSerIleThrGlnGluAlaMetGly----- 283  
 QY 2106 TCCTGACTGGGAGACATCTCCCATTTAGGGCCCAATAGACATTTTCATACAGGAGACAGG 2165  
 DB 284 ----- 284  
 QY 2166 CTGGAGTGGACCTCCAGCAAACTCCAGCAGACACCTGCAGCAGAGCGGCGCTGACTGTGTA 2225  
 DB 284 eTrpValGlyGln----- 288

```
QY 2226 GGAAGTAAACACAGAAAGGAATAGTATCAAC-----ATTAACA 2267
Db      |||:::||||:::
289 -----GlnArgLysArgGlnProLeuGlyTyrGluGluArgLysLeuThrAsnAr 305
QY 2268 AAAGGACATCCACTCAGAGACCCCTCTGAGGTCACAAACATCAAGACCAAAAGGTAA 2327
Db      :|||||:::|||||:::|||||:::|||||:::
305 glysAspIleHisIleLysAsnProSerValcysHisHisGlnArgProLysValas 325
QY 2328 TAAACCAAAAAGATGGGAAAAAC-CAGTCGACGAACACTGAAATTCACAAAACACGAA 2386
Db      |||:::|||||:::|||||:::|||||:::|||||:::
325 pLysThrThrLysMetGlyLysLysGlnSerArgLysThrGlyAsnSerLysAsnGlnSe 345
QY 2387 CTCCTCTTCTCAACAAAGATCAACACTCTCCGACGAAGGAACAAACACAGATGA 2446
Db      :|||:::|||||:::|||||:::|||||:::|||||:::
345 rThrSerProThrLeuLysGluSerSerSerProAlaThrGluGlnSerTprThrGl 365
QY 2447 CAATGAGTTTTCAGGAATTCAGAGATGAGGCTTCAGAAGGTGGGTAAATAACAACTCCTC 2506
Db      |||:::|||||:::|||||:::|||||:::|||||:::
365 uAsnAspPheAspGluLeuArgGluGluGlyPheArgArg-----SerAspTyrSe 382
QY 2507 CGAGCTAAAGGAGATGTTCTAACCCCAATGCAAGGAAGCTTAAGAACCTTGAAGAAAGCTT 2566
Db      |||:::|||||:::|||||:::|||||:::|||||:::
382 rGluLeuGlnGluGluValGlnAsnAsnGlyLysGluValLysAsnPheGlyLysLysLe 402
QY 2567 AGATCAATTCAGTAATACTAGATGAGTGTAGAGAAGAACATAAATGACCTGATGGAGCT 2626
Db      |||:::|||||:::|||||:::|||||:::|||||:::
402 uAspGluThrIleThrArgIleThrAsnAlaLysLysSerLeuLysAspLeuMetGluLe 422
QY 2627 GAAAACCCAGACAGAACACTTCATGAAGCATACACAGCTTCATAGCCCAATCGATCA 2686
Db      |||:::|||||:::|||||:::|||||:::|||||:::
422 uLysThrLysAlaArgGluLeuLysAspGluArgThrSerLeuSerSerGlnCysAsnGl 442
QY 2687 AGCAGAAGAAAGGATATCAGTGATGAAGATCAATTAATAAAGAAAGTGAAGAGACAA 2746
Db      |||:::|||||:::|||||:::|||||:::|||||:::
442 nLeuGluGluArgValSerValMetGluAspGluMetAsnGluLysGlnGluGluLys 462
QY 2747 GATTACAGAAAAGAGTGAAGAAACAAACAAAGCTCCCAAGATTAATGGGACTATGT 2806
Db      |||:::|||||:::|||||:::|||||:::|||||:::
462 sPheArgGluLysArgIleLysArgAsnGluGlnSerLeuGlnGluIleTprAspTyrMe 482
QY 2807 GAAAGACCAATCTACATTTGATGTGGTGCTCCCAAGTATGGGAGATGAATCA 2866
Db      :|||||:::|||||:::|||||:::|||||:::|||||:::
482 tLysArgProAsnLeuArgLeuIleGlyValProGluSerAspGlyGluAsnGlyThrLy 502
QY 2867 GTTGAAACACTCTCTCAGGCTATATCCAGGAGAAATTTCCCAT-CTATCAGGCGAGCC 2925
Db      |||:::|||||:::|||||:::|||||:::|||||:::
502 sLeuGluAsnThrLeuGlnAspIleGlnGluAsnPheProAsnLeuAlaArgGlnAl 522
QY 2926 CAACATTCAAATTCAGGAATATGGAGAACACCATTAAGATACTCTCGAGAGAACAAT 2985
Db      |||:::|||||:::|||||:::|||||:::|||||:::
522 aAsnIleGlnIleGlnGluGlnArgThrProGlnArgTyrSerLeuArgArgAlaTh 542
QY 2986 CCCAAGACACATAATCTTCAGATTCACCAAGGTGAATGAAGGAAAGAAATGTTAAGGC 3045
Db      |||:::|||||:::|||||:::|||||:::|||||:::
542 rProArgHisIleIleValArgPheThrLysValGluMetLysGluLysMetLeuArgva 562
QY 3046 ACCGAGAGAAAAGGTTGGGTACCCCAAGGAAGGCAATACAGACTTAACAGCGGATCT 3105
Db      |||:::|||||:::|||||:::|||||:::|||||:::
562 lAlaArgGluLysGlyArgValThrHisLysGlyLysProValArgLeuThrAlaAspLe 582
QY 3106 CCCGGCAGAAACCTTACAGCCAGAGAGATGAGGGCCCAATATTCACATCTCTTAAGA 3165
Db      |||:::|||||:::|||||:::|||||:::|||||:::
582 uLeuAlaGluThrLeuGlnAlaArgArgGln----- 592
QY 3166 AAATAAATTTTCAACCCAGAAATTTTCATATCCAGCCAAACCAAGCTTCCTAAGTGAAGGAGA 3225
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3226 AATAAAATCTCTTACAGAGAAGCAAAATGCTGACAGATTTTGTCCACCAGCGCTGCCT 3285
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3286 TACAAGAGCTCTCTGAAGGAAGCCACCAACATGGAAGGAACAACTGGGTACCGACCTGCA 3345
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3346 AAAACATCCCAAAATGTTAAAGACCATTCATGCTATGAAGAAAGTGCATCAACTAACGGGC 3405
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3406 AAAATAACAGCTAGTGTGTATTAATGCGCAGGATCAAAATTCACACATAAATAATTAACTT 3465
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3466 AAATGTAATGGGTAAATTCCTCCCAATTAAGAACACAGACTGGCAAAATTCGATAAAGAG 3525
Db      |||:::|||||:::|||||:::|||||:::|||||:::
592 ----- 592
QY 3526 TCAAGACCCATCAGTGTGCTCTATTTCAGGAGGCCCATCTCACATGAAAAGACACACATAG 3585
Db      |||:::|||||:::|||||:::|||||:::|||||:::
593 -----AspThrHisAr 596
QY 3586 GCTCAAAATAAAGGATGGAGAGATTTTACCAGTAATTAAGAAAACAAAAAAGC 3645
Db      |||:::|||||:::|||||:::|||||:::|||||:::
596 gLeuLysIleLysGlyTprArgLysIleTyrGlnAlaAsnGlyLysGln---LysLysAl 615
QY 3646 AGGGTTGCAATCTCTCTCTGTATATAAAGACAGCTTTAAACCAACAAAGATCAAAAGAGA 3705
Db      |||:::|||||:::|||||:::|||||:::|||||:::
615 aGlyValAlaIleLeuValSerAspLysThrGluPheLysProThrLysIleLysArgAs 635
QY 3706 CAAAGAGGCCATACATAATGGTAAAGGCATCAATGAACAAAGAGAGCTAACTATCTCT 3765
Db      |||:::|||||:::|||||:::|||||:::|||||:::
635 pLysGluGlyHisTyrIleMetValLysGlySerIleGlnGlnGluGluLeuThrIleLe 655
QY 3766 AAATATACATCACCCCAATACAGGACCCAGATTCATAAGCAAGTCTCTTAGAGACCT 3825
Db      |||:::|||||:::|||||:::|||||:::|||||:::
655 uAsnIleTyrAlaProAsnThrGlyAlaProArgPheThrLysGlnValLeuLeuArgAspLe 675
QY 3826 ACAAGAGACCTTCTACTCCCAACACATAATAGTGGGAGCTTAATAATAAATAGACACTT 3885
Db      |||:::|||||:::|||||:::|||||:::|||||:::
675 uGlnArgAspLeuAspSerAsnThrIleIleThrGly-----AspPh 689
QY 3886 TAACACCCCTGCGCAATATTAGGACATCAATGAGACAGAAAAATTAACAGGATATCCA 3945
Db      |||:::|||||:::|||||:::|||||:::|||||:::
689 eAsnThrProLeuSerThrLeuAspArgSerMetArgGlnLysValAsnLysAspIleGl 709
QY 3946 GGAGTTGAACCTGAGCTCTGGACCAAGCGGACCTAATAGATATCTACAGACTCCCCACC 4005
Db      |||:::|||||:::|||||:::|||||:::|||||:::
709 nGluLeuAsnSerAlaLeuHisGlnAlaAspLeuIleAspIleTyrArgThrLeuHisPr 729
QY 4006 CAATCAACAGAAATATACACTCTTCTCAGCATCATTTACACCTATTTTAAATATGACCA 4065
Db      |||:::|||||:::|||||:::|||||:::|||||:::
729 oLysSerThrGluAsnThrPhePheSerAlaProHisHisThrTyrSerLysIleAspHI 749
QY 4066 TGTAAATTTTAAGTAAACACTCTCTCAGCAAAATGCAAAAGACAGAAATCTTAACAAACAG 4125
Db      |||:::|||||:::|||||:::|||||:::|||||:::
749 sIleValGlySerLysAlaLeuLeuSerLysCysLysArgLysGluIleIleThrAsnCy 769
QY 4126 TCTCTCAGACTACAGTCAATCTATTTAGAACTCAGAAATTAAGAACTCACTCAAAATCA 4185
Db      |||:::|||||:::|||||:::|||||:::|||||:::
769 sLeuSerAspHisSerAlaIleLysLeuGluLeuArgIleLysLysLeuProGlnAsnCy 789
QY 4186 CACAACATACATGGAACCTGCTCTCTCAATCACTACTCTGGTAAATAACAAAT 4245
Db      |||:::|||||:::|||||:::|||||:::|||||:::
789 sSerThrIleThrLysLeuAsnAsnLeuLeuLeuAsnAspTyrTrpValHisAsnGluMe 809
QY 4246 GAAGGCAAAAATAAAGATGTTCTTTGAAACCAATGAGAACCAAGACACAAATGTACCAGAA 4305
Db      |||:::|||||:::|||||:::|||||:::|||||:::
809 tLysAlaGluIleLysMetLeuPheGluThrAsnGluAsnLysAspThrThrTyrGlnAs 829
QY 4306 TCTCTGGGCGCATATTTAAAGCAGTGTGTAGAGGAAATTTATAGACTAGTAGCTACAA 4365
Db      |||:::|||||:::|||||:::|||||:::|||||:::
829 nLeuTrpAspThrLeuLysAlaValCysArgGlyLysPheIleAlaLeuAsnAlaHisGl 849
QY 4366 GAGAAACGAGAAATATCTAAATATAGACACCTTAACTCACAATTAAGAACTAGAGAA 4425
Db      |||:::|||||:::|||||:::|||||:::|||||:::
4425 ----- 4425
```

Db 849 yArgLysGlnArgSerLysIleAspThrLeuThrSerGlnLeuLysGluLeuAlaLys 869  
Qy 4426 GAAAGACAAACAAATTCAAAGCTAGCAGAGACAGAAATTAACTAAGATCAGAGCAGA 4485  
Db 869 sGlnGluGlnThrHisSerLysAlaSerArgArgGlnGluIleThrLysIleArgAlaG 889  
Qy 4486 ACTGAGGAGATGAGACACAAAGCCCTTCAAATTAATCAATGAATCCAGGAGCTGGT 4545  
Db 889 uLeuLysGluIleGluThrGlnLysThrLeuGlnLys-IleAsnGluSerArgSerTrpP 909  
Qy 4546 TTTTGAAGATCAGCAAAATAGACCACCTAGACAGACTAATAAGAGAGAGAGAGAA 4605  
Db 909 he-----Leu--ProLeuAlaArgLeuIleLysLysThrGluLys 922  
Qy 4606 GAATCAAGAGATGCAATAAAATGATAAGGGGATATCACCACCGATCCACAGAAAT 4665  
Db 922 sAsnGlnIleAspAlaIleLysAsnAspLysGlyAspIleThrAsnProThrAspII 942  
Qy 4666 ACAAACTATTATCAGAGAATATTATAACACCTCTATGCAAAATAAACTAGAAAATCTAGA 4725  
Db 942 eGlnThrAlaIleArgGluTyrLysHisLeuTyrAlaAsnLysLeuGluAsnLeuG 962  
Qy 4726 AGAAATGGAATAATCTCTGACACATATGTAGCCTGTATGGACCTTGGGGGACAGACAA 4785  
Db 962 uGluMetAspLysLeuLeu 968  
Qy 4786 AAGGGGTGAATGCAGAAATAAAAGACAAAGAGAGTATGTTTGAAGTAGGGGTC 4845  
Db 968 ----- 968  
Qy 4846 AGGGGCACTTCGCTTAATGACAAAGGCCCTGAGCTTTACACCACCCCTCTGTATTTA 4905  
Db 968 ----- 968  
Qy 4906 TTAGGCAAAAGAGATAGCAGAGGGTGAGTTGGAAGAAGAGTCAGCTGTTAGGTCAGA 4965  
Db 968 ----- 968  
Qy 4966 GTAGGCTGCAAGACTGCATTCTCAACAATAGGCTCTAGATGTCCAGTAGATAACCT 5025  
Db 968 ----- 968  
Qy 5026 CAAGGACCAAGTCCAGGGAGTATGGCCCTCAGCAAAACCTTCTAGGGCAGGCACAGAG 5085  
Db 968 ----- 968  
Qy 5086 TAAGTTGGCCACATCTGTATTCAGGATAACAGTTGCTGTTGTATCAAGTAGCCTCC 5145  
Db 968 ----- 968  
Qy 5146 AGTGAATGCTGAGTTGGTCATGATCCCTTTGGCCTTTGGCTCCCAAAACACATACAC 5205  
Db 969 -----AsnThrTyrTh 972  
Qy 5206 CCTCTCAAGACTAAACAGGAGAGTCAAAATCCCTGAATATACAGTAACAAGTCTTAA 5265  
Db 972 rIleProArgLeuAsnGlnGluValGluProArgAsnArgProIleThrGlySerG 992  
Qy 5266 AATTGAGCAGTAATTGATAGCCTACCAACCAAAAAGTCCAGGACGAGGATTCAC 5325  
Db 992 uIleGluAlaIleAsnArgLeuProThrLysLysCysProGlyProAspGlyPheTh 1012  
Qy 5326 AGCAAAATCTACAGAGGTACAAAGAGAGCTGGTACTATTCTCTGAAACTATTCCA 5385  
Db 1012 rAlaGluPheTyrGlnArgTyrLysGln-----ValProPheLeuLysLeuPheG 1030  
Qy 5386 AAAATAGAA---AATGGGAATCCCTCAACTCATTTTACGAGGCGCAGCATCCTGA 5442  
Db 1030 nSerThrGluLysGly-GlyIleLeuProLysSerPheTyrGluAlaSerIleIleu 1050  
Qy 5443 TACCAAACTGACAGTGACACAAAGAGAGAAATTTACGGCCCATATCCCTGATGA 5502  
Db 1050 leProLysProGlyLysAspThrThrLysLysGluAsnPheArgProIleSerLeuMeta 1070

Qy 5503 ACATTGTGTGAAATCCTCAATAAATACTGCAAAACCAAAATCCAGCAGCACATCAAAA 5562  
Db 1070 snIleAspValLysIleLeuAsnLysIleLeuAlaLysGlnIleGlnHisIleLysL 1090  
Qy 5563 AGTTTATCTACCATGATCAAGTTGGGCTCATCCTCGGATGCAAGGCTGTTCAAAATAT 5622  
Db 1090 ysLeuIleHisAspGlnValGlyPheIleProGlyMetGlnArgTrpPheAsnIleA 1110  
Qy 5623 GCAAAATCAATAATGTAGGCCATCACATAAACAGAACCAATGACAAAAACCATGATTA 5682  
Db 1110 rGlySerIleAsnValIleGlnHisIleAsnArgThrLysAspLysAsnArgMetIleI 1130  
Qy 5683 TCTCAATAGATGCAGAAAAGCCCTTGTCAAAATTCAAACAGCCCTTCATCCTAAAAATTC 5742  
Db 1130 leSerIleAspAlaGluLysAlaPheAspLysIleGlnArgPheMetLeuLysThrL 1150  
Qy 5743 TCAGTAACTAGTATCGATGGAATGTATCTCAAAATAATAAGAGCTATTTATAC-AAAC 5801  
Db 1150 euAsnLysLeuGlyIleAspGlyThrTyrLeuLysThrIleArgAlaIleTyrAspLys 1170  
Qy 5802 CCACAGCCATATCATCTCAATGGGCAAAACTGGAGCATTCCTTTTGAGAACTGGCA 5861  
Db 1170 roThrAlaAsnIleIleLeuAsnGlyGlnLysLeuGluAlaPheProLeuLysThrGlyT 1190  
Qy 5862 CAAGCAAGGATGCCCTCTCTCACCACTCCTATTCAAGATPACTATTGGAAGTTCTGGCCA 5921  
Db 1190 hrArgGlnGlyCysProLeuSerProLeuLeuPheAsnMetValLeuGluValLeuAla 1210  
Qy 5922 GGCAATCAGGCAATAGAAAGAAATAAGGGTATTCATAAGAGAGAGAGCAAGTCATAT 5981  
Db 1210 rAlaIleSerGlnGluLysGluIleLysAspIleGlnLeuGlyLysGluGluValLysL 1230  
Qy 5982 TGTCTCTGTTGCAAGTACATGTTTGTATATTAGAAAACCCATCGTCTCAGGCCAAA 6041  
Db 1230 euSerLeuPheAlaAspAspMetIleValTyrLeuGluAsnProIleValSerAlaGlnA 1250  
Qy 6042 AACTCCTTAAGCTGATAAGCAACTCTCAGCAAGCTCTCAGGACACAAAATCAATGTGCAA 6101  
Db 1250 snArgLeuLysLeuIleSerAsnPheArgLysValSerGlyTyrLysIleAsnValGlnL 1270  
Qy 6102 AATCACAAGCATCTTATATACCCCAATTAATAGACAACAGAGAGCCCAATCATGAGTAAC 6161  
Db 1270 ysSerGlnAlaPheLeuTyrThrAsnAsnSerGlnThrGluSerGlnIleMetSerGluL 1290  
Qy 6162 TCTCATTCACAATTGCTACAAAGACAGATAAATACCTAGCAATACACTTACAGGAGACA 6221  
Db 1290 euProPheThrAlaSerLysArgIleLysTyrLeuGlyIleGlnLeuThrArgAspV 1310  
Qy 6222 CGTAGGAACCTCTCAAGGAGAACTACAACCACTGATCAAGGAAATAAGAGAGAGACACA 6281  
Db 1310 allysAspLeuPheLysGluAsnTyrLysProLeuLeuSerGluIleLysGluAspThra 1330  
Qy 6282 ACAATGGAAAAACATCTCCATGCTCACAGATA-----GTAGAATCATGAAAAATGCC-A 6334  
Db 1330 snLysCysLysAsnValProCysSerTyrValGlyArgIleAsnIleMetLysMetaI 1350  
Qy 6335 TACTGCCCAAGTAATATTATAGATTAGTCTACCCCACTCAAGCTACCATTCACCTTCT 6394  
Db 1350 leLeuProLysValIleTyrArgPheAsnAlaIleProIleLysLeuProMetThrPheP 1370  
Qy 6395 TCACAGAAATGGAAAAACAACCTTAAATTTTCATATGGAAACCAAAAAAGAGCCACAGA 6454  
Db 1370 heThrGluLeuGluLysThrThrLeuLysPheIleIleIleIleIleIleIleIle 1389  
Qy 6455 GCAAGACAACTTTAAGCAAAAAAGAACAAAGCTGGAGGTATCATGCTACTGACTTAAAA 6514  
Db 1390 AlaLysSerIleLeuSerGlnLysAsnLysAlaGlyIleThrLeuProAspPheLys 1409  
Qy 6515 CTATACATTAAGCTACAGTAACCAAACTGCATGCTGTTACTGTTACCAAAACAGATATATA 6574  
Db 1410 LeuTyrTyrLysAlaThrValThrLysThrThrTrpTyrTrpTyrGlnHisArgAlaVal 1429

